

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 19:09:19 ; Search time 3550 Seconds

(without alignments)

10659.738 Million cell updates/sec

Title: US-10-029-180-7

Perfect score: 1557

Sequence: 1 atggatcctagaaccatcc.....ccgttcagccgacgacaa 1557

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vit.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_nus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	380.4	24.4	442	9	AA785001 g4b03a1.r
2	171.6	11.0	768	12	EM077105 TREST-A34
3	108.8	7.0	687	10	BF251517 EST418778
4	82.8	5.3	782	28	AQ161501 mgxb0008A

5	82	5.3	585	28	AQ255041
6	56.2	3.6	807	13	AX464554
7	53.8	3.5	1448	29	CC220110
8	52.6	3.4	1201	9	AL515475
9	52.2	3.4	727	13	AX319543
10	51.4	3.3	1137	29	CC315110
11	50.6	3.2	567	28	AZ895651
12	50	3.2	1201	13	AX457568
13	49.6	3.2	1201	13	AX421672
14	49.4	3.2	802	29	CNS01110
15	49.2	3.2	786	28	AQ745021
16	49.2	3.2	897	29	AG060530
17	49.2	3.2	915	13	BO952432
18	49.2	3.2	1262	12	BM560943
19	49.2	3.2	1287	10	BF256624
20	49	3.1	481	9	AA695157
21	49	3.1	759	13	BU314584
22	48.8	3.1	834	28	BZ051898
23	48.8	3.1	1101	29	CNS00008
24	48.8	3.1	1201	9	AL514195
25	48.6	3.1	458	10	BF063108
26	48.6	3.1	648	12	B1227246
27	48.6	3.1	1026	29	AG081192
28	48.4	3.1	732	29	EX149729
29	48.4	3.1	815	29	EX239144
30	48.4	3.1	987	29	CNS00418
31	48.4	3.1	1137	29	AG078502
32	48.2	3.1	786	28	BH018846
33	48.2	3.1	2152	29	CC283954
34	48	3.1	785	29	AG081505
35	48	3.1	1130	29	AG060387
36	48	3.1	1201	13	AX356664
37	47.8	3.1	759	29	AG126157
38	47.8	3.1	774	29	AG128958
39	47.8	3.1	887	29	CNS02SC6
40	47.8	3.1	1082	9	AL574528
41	47.8	3.1	1104	13	BO671688
42	47.8	3.1	1122	29	CC231036
43	47.6	3.1	731	29	CNS04208
44	47.6	3.1	789	14	CA320647
45	47.6	3.1	1201	13	AX365323

ALIGNMENTS

RESULT 1
AA785001
LOCUS
DEFINITION
442 bp mRNA linear EST 29-JUL-1998
g4b03a1.r1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans cDNA clone
g4b03a1 5', mRNA sequence.
AA785001
AA785001.1 GI:2845169
EST.
SOURCE
ORGANISM
Emericella nidulans (anamorph: Aspergillus nidulans)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
REFERENCE
AUTHORS
Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R., Prade, R. and Roe, B.
1 (bases 1 to 442)
TITLE
JOURNAL
COMMENT
An Aspergillus nidulans EST Database
Unpublished
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center


```

674 ATCCGGCGGGCAACCCCAAGGCGCGCTTGGCCCTTTGGTCCC 717
|||||
Db

RESULT 3
BF251517
LOCUS
DEFINITION
EST418778 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAA91 5' sequence, mRNA sequence.
ACCESSION
BF251517
VERSION
BF251517.1 GI:16931360
KEYWORDS
EST.
SOURCE
Coccidioides immitis
Coccidioides immitis
ORGANISM
Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
1 (bases 1 to 687)
AUTHORS
Gardner,M.J. and Kirkland,T.
TITLE
Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL
Unpublished
COMMENT
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.

FEATURES
Location/Qualifiers
source
1..687
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAA91"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/notes="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
Xho1"

BASE COUNT 167 a 234 c 160 g 126 t
ORIGIN

Query Match 7.0%; Score 108.8; DB 10; Length 687;
Best Local Similarity 56.2%; Pred. No. 9.3e-17;
Matches 396; Conservative 0; Mismatches 252; Indels 56; Gaps 8;

QY 95 TGCCTCAGTACAGATGAGCTCAGTACCGCTCTCTCAGCGGACACCCCTGCTCTC 154
Db 30 TCCACCAATATCAGATGCAATTCGCCGTACGCGGCAACCGCATAGTTGCCGCCAT 89

QY 155 TGCACCCCATATAGCCAGTCCGCGCTCTCTCACTCGTACATGGGCGGCGCGTACC 214
Db 90 TGCACACCATCAGATCGTCCGCAATTCAGCATAGTATCTC-----CAGCCTTTCC 143

QY 215 GGCCTGATCTGAACAGGTACCCCGCA---TCAAGTCAAGATTTACGGCTTCTCTGCTG 271
Db 144 GGAATGACATGCTAGATACCCGCAACCTCCGCGACCGATGTTTATGCGGTATCGTCGG 203

QY 272 GCGCATATGCCCACACTACGTTGGGAGCTTGCCTCCGACATCTTCTCTTCATC 331
Db 204 CCGCGTTGACTACGATGCGCGCGTCAACAGTTTGCACCGCTCACTTTCTT----- 256

QY 332 CCAATCCGCGGCGGACGAGGCGGACGCAATCGCGCATCTCTCTCTCATAGCG 391
Db 257 -----GGCCACCAATCCACGCAATTCAGCATATCA 293

QY 392 TGCTCCGCGCGCTTCCAGCGCTCAGTGTATCCCGAGCCAAATGCGCGGCGCTCCCC 451
Db 294 TGCTCCCAACCCAGCAGCAGCCACCAAGCATATCCCGAACCAATCGCGCGGCGCACCGC 353

QY 452 GGGACCGTCTGCTGACTTCAATGACTTCTCTTCAGGACATTCAGTATTCGAGC 511
Db 354 GTGACAGAAGATCAGAGTATGCGCTG---TGCGGTGGCGCCCGTTTCAAAACGGGAGA 410

us-10-029-180-7.rst
|||||
QY 512 GAAAGCTTCAAGGTTGGGACCCCGTTGCTG-----CGAATGGTGTCTGCGCGTATCCCG 565
Db 411 ACAAGCAACCCGTATGGCTGGCGCAGAAGACTGCCACCCACCTCTGGCCCTTACATGC 470

QY 566 GGAAGACTCCCGCCGAACCCAGGTTGTGTTCTCAGGGGGGAGCGCGTATCTCTCCGA 625
Db 471 CAAAGATCTCTCCGAGGACACAAGTGTAGGAACCCAGGCGCGGAGAGTATCTCTCCGA 530

QY 626 GTGTTCCGGGACGCGCACTCCGGTTCACAATGGTGTAAACGGCACCGGCAAGACACTA 685
Db 531 GTGTACCGGGTCCGCTGC-----GTTTACAACGGTGTGAATGGCAC-----CAGGGGAACA 583

QY 686 CTATCCCGGCCAAGGATGCCGA-TGGAAAGTTCCTTTCGCCGGAATGTAAACAAGACTTAT 744
Db 584 CCATCCCTGCCAAGGATGCAGACGGGAAGTTTCCTGTCCGCACTGCAATAAGACTTTA 643

QY 745 CTTATGCCAAGCATCTCAAGCGCATCTGTAGCCACACTGG 788
Db 644 CTGGCATGCGAAGCACCTTAAAGCGCATCTCTTAGCCACACCGG 687

RESULT 4
AQ161501
LOCUS
DEFINITION
clone mgxb0008A02r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
clone mgxb0008A02r, genomic survey sequence.
ACCESSION
AQ161501
VERSION
AQ161501.1 GI:3557902
KEYWORDS
GSS.
SOURCE
Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM
Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
1 (bases 1 to 782)
AUTHORS
Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
TITLE
A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome
JOURNAL
Unpublished
COMMENT
Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdeane@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 315.
Location/Qualifiers
1..782
/organism="Magnaporthe grisea"
/mol_type="genomic DNA"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgxb0008A02r"
/tissue_type="Protoplasts"
/lab_host="B. coli DH10B"
/clone_lib="CUGI Rice Blast BAC Library"
/notes="Vector: pBACWICH; Site 1: HindIII; Site 2: HindIII;
Rice blast is one of the most devastating fungal diseases
of rice world wide. It is a filamentous ascomycete with
a haploid genome (n=7) of approximately 40 Mbp. Rice
blast is an important model fungal pathogen for studying
numerous aspects of the fungal-host interaction. In
order to facilitate genome wide analysis, a BAC library
containing 9216 clones with an average insert size of 130
kbp was constructed. This library represents greater
than 25X genome coverage. High density colony filters
are available upon request."
BASE COUNT 172 a 284 c 192 g 134 t
ORIGIN

```

Query Match 5.3%; Score 82.8; DB 28; Length 782;
Best Local Similarity 51.5%; Pred. No. 3.8e-10;
Matches 271; Conservative 0; Mismatches 242; Indels 13; Gaps 3;

QY 321 CCTTCTCATCCCAATCCGAGCGGAGGACGAGGCGGAGCAATCGCGCACTATCCTCC 380
DB 116 CCGTATACCTGGCATGATGCCGCGCGGAGTACCAACCACTTACCCCGGCTATCCAC 175
QY 381 TCCTCATAGCGTCTCCGCGCGCTTCCAGCGCTCAGTGGTATCCCGCAGCCCAATTCGCGC 440
DB 176 CACAACCTCAATGATGACCCAGGATCGAGCGCAATGGCCCACTCACTTATTCGCGC 235
QY 441 GGGCGCTCCCGGAGCGTGTCTGACTTCAACATGAGCTTCTTCAGAGCATTCAG 500
DB 236 CGGCGCTCTCTGT 295
QY 501 TTATTTCGAGCGGAAG-----CCTCAAGGTTGGAGCCCGTGTCTCGAATGTGTGTGT 555
DB 296 CCTCAGGCGGCGATGT 353
QY 556 CCGTATCCCGGGAAGGACTCCCGCGAACCAGGTTGTGTGTGTGTGTGTGTGTGTGTGT 615
DB 354 ATGATCAACAGGAGGAGCGGAGCGGACACATGTGTGTGTGTGTGTGTGTGTGTGTGT 413
QY 616 ATCTCTCCGAGTGTTCGGAGCGGCACTCCCGGTCAAAATGTGTGTGTGTGTGTGTGT 675
DB 414 ATCTCTCCGAGCGCGCGGTAGG-----CCCCAGCGCGCGTGTGTGTGTGTGTGTGT 467
QY 676 AAGACACTACTATCCGCGCAAGGATCGCGGATCGCGGATCGCGGATCGCGGATCGCG 735
DB 468 GCAAGAACAGATCCCGGAGAGGATCGCGGAGGATCGCGGAGGATCGCGGAGGATCG 527
QY 736 AAGACTTATCTTCATGCAAGCACTCAAGCGGCATCTGCTAGCCACATCTGTGTGTGTGT 795
DB 528 AAGAGTACTCTGACGCGCAAGCATCTCAAGAGGATCGCGGAGGATCGCGGAGGATCG 587
QY 796 CCGTACATGT 841
DB 588 GCTTTGT 633

RESULT 5
AQ255041 585 bp DNA linear GSS 23-OCT-1998
LOCUS mgxb0008J11r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
DEFINITION clone mgxb0008J11r, genomic survey sequence.
ACCESSION AQ255041
VERSION AQ255041.1 GI:3779356
KEYWORDS GSS.
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Magnaporthe grisea
REFERENCE 1 (bases 1 to 585)
AUTHORS Yu, Y.; Zhu, H.; Boyd, C.A.; Gaudette, B.; Gayle, A.; Kingsbury, R.;
Phillips, K.; Sasinowski, M.; Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome
JOURNAL Unpublished
COMMENT Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 236.
Location/Qualifiers
1. 585
/organism="Magnaporthe grisea"
/mol_type="genomic DNA"

/strain="70-15"
/db_xref="taxon:148305"
/clone="mgxb0008J11r"
/tissue type="Protoplasts"
/lab host="E. coli DH10B"
/clone lib="CUGI Rice Blast BAC Library"
note="Vector: pBACWICH; Site 1: HindIII; Site 2: HindIII;
Rice blast is one of the most devastating fungal diseases
of rice world wide. It is a filamentous ascomycete with
a haploid genome (n=7) of approximately 40 Mbp. Rice
blast is an important model fungal pathogen for studying
numerous aspects of the fungal-host interaction. In
order to facilitate genome wide analysis, a BAC library
containing 9216 clones with an average insert size of 130
kbp was constructed. This library represents greater
than 25X genome coverage. High density colony filters
are available upon request." 93 t

BASE COUNT 127 a 220 c 145 g 93 t
ORIGIN

Query Match 5.3%; Score 82; DB 28; Length 585;
Best Local Similarity 53.2%; Pred. No. 5.3e-10;
Matches 248; Conservative 0; Mismatches 205; Indels 13; Gaps 3;

QY 321 CCTTCTCATCCCAATCCGAGCGGAGGACGAGGCGGAGCAATCGCGCACTATCCTCC 380
DB 116 CCGTATACCTGGCATGATGCCGCGCGGAGTACCAACCACTTACCCCGGCTATCCAC 175
QY 381 TCCTCATAGCGTCTCCGCGCGCTTCCAGCGCTCAGTGGTATCCCGCAGCCCAATTCGCGC 440
DB 176 CACAACCTCAATGATGACCCAGGATCGAGCGCAATGGCCCACTCACTTATTCGCGC 235
QY 441 GGGCGCTCCCGGAGCGT 500
DB 236 CGGCGCTCTCTGT 295
QY 501 TTATTTCGAGCGGAAG-----CCTCAAGGTTGGAGCCCGTGTGTGTGTGTGTGTGTGT 555
DB 296 CCTCAGGCGGCGATGT 353
QY 556 CCGTATCCCGGGAAGGACTCCCGCGAACCAGGTTGTGTGTGTGTGTGTGTGTGTGTGT 615
DB 354 ATGATCAACAGGAGGAGCGGAGCGGACACATGTGTGTGTGTGTGTGTGTGTGTGTGT 413
QY 616 ATCTCTCCGAGTGTTCGGAGCGGCACTCCCGGTCAAAATGTGTGTGTGTGTGTGTGT 675
DB 414 ATCTCTCCGAGCGCGCGGTAGG-----CCCCAGCGCGCGTGTGTGTGTGTGTGTGT 467
QY 676 AAGACACTACTATCCGCGCAAGGATCGCGGATCGCGGATCGCGGATCGCGGATCGCG 735
DB 468 GCAAGAACAGATCCCGGAGAGGATCGCGGAGGATCGCGGAGGATCGCGGAGGATCG 527
QY 736 AAGACTTATCTTCATGCAAGCACTCAAGCGGCATCTCAAGCGGCATCTGTGTGTGTGT 781
DB 528 AAGAGTACTCTGACGCGCAAGCATCTCAAGAGGATCTCAAGAGGATCTCTGTGTGTGT 573

RESULT 6
BX464554 807 bp mRNA linear EST 22-MAY-2003
LOCUS BX464554 Homo sapiens PLACENTA Homo sapiens cDNA clone CSODE007YB11
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX464554
VERSION BX464554.1 GI:31021646
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 807)
AUTHORS Li, W.B.; Gruber, C.; Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished

COMMENT

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqrefgenoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3370.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS1DE0022F03NP1&cluster=3370.r. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS1DE0022F03NP1.
Location/Qualifiers
1. 807
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE007XB11"
/tissue type="PLACENTA"
/clone lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
5 a 609 c 15 g 16 t 162 others

FEATURES

source

BASE COUNT

ORIGIN

Query Match 3.6%; Score 56.2; DB 13; Length 807;
Best Local Similarity 31.5%; Pred. No. 0.0021;
Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;

Qy 15 CCATCCCTCTCGGCTCGTCTACAGTCTGCTCAAGGATCGGCGCTCTTCCTTCGC 74
Db 230 CCC 289

Qy 75 TCCATCTCGAGCATGCGCAATGCTCTAGTACAGATGAGCTCTAGTACCCAGTCTCTCA 134
Db 290 CCC 349

Qy 135 GCGCAGACCTGCTCTCTGCAACCCCATATAGCAGTGCCTGCTCTCTCTCTGTA 194
Db 350 CCC 409

Qy 195 CATGGGCGAGCGCGTACCGGCTGATCTGAACAGGTACCCGATCATGCTCAGATGT 254
Db 410 SCC 469

Qy 255 TTACGGGTCTTCTGTCGCGCGATAATGCCCACTACCTACCGTGGGAGCTTGCCTCCGAC 314
Db 470 CCC 529

Qy 315 ATCTTTCTTCTATCCCAATCGCAGGCGAGGACAGCGGAGCAATCGCGCACTA 374
Db 530 CCC 589

Qy 375 TCCTCTCTCTCATAGCGTGTCCCGCGCTTCCAGCGCTCAGTCTACCGCAGCCAA 434
Db 590 CCC 649

Qy 435 TGCCTCCGCGCTCCCGGAGCGCTGCTGTGACTTCAACAAATGAGCTTCTCTCAGGAGC 494
Db 650 YCCCYCCYKCCYCGBBKYYTCYBKSTGSKYKGSYBCKKSSSSSSCC 709

Qy 495 ATTCAATTATCGACGAAGCTCAAGTTGGGACCCCTTCTGCGAATGCTGTCG 554
Db 710 CTYSBKCKKTBKGGYTYBTSCCYCCBYCYCTCSBKGSCKGSKBSYBYCB 769

Qy 555 GCCGTATCCCGGAGGACTCCCC 579
Db 770 YBCYYKSTGTCTCGCSYKCCSY 794

RESULT 7

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 1448

Location/Qualifiers

/organism="Gallus gallus"

/mol_type="genomic DNA"

/strain="Red Jungle Fowl"

/db_xref="taxon:9031"

/clone="CH261-92F21"

/sex="female"

/cell_line="UCD001, inbred 256"

/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;

CH261 Female Chicken library - for library and clone

ordering information: http://www.chori.org/bacpac"

BASE COUNT 32 a 82 c 1229 g 33 t 72 others

ORIGIN

Query Match 3.5%; Score 53.8; DB 29; Length 1448;

Best Local Similarity 45.6%; Pred. No. 0.011;

Matches 190; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

Qy 15 CCATCCCTCTCGGCTCTCGTCTACAGTCTGCTCAAGGATCGGCGCTCTTCCTTCGC 74

Db 806 CCC 747

Qy 75 TCCATCTCGAGCATGCGCAATGCTCTAGTACAGATGAGCTCTAGTACCCAGTCTCTCA 134

Db 746 CCC 687

Qy 135 GCGCAGACCTGCTCTCTGCAACCCCATATAGCAGTGCCTGCTCTCTCTCTGTA 194

Db 686 CCC 627

Qy 195 CATGGGCGAGCGCGCTACCGGCTGATCTGAACAGGTACCCGATCATGCTCAGATGT 254

Db 626 CAGGGGCGCGGCCC 567

Qy 255 TTACGGGTCTTCTGTCGCGCGATAATGCCCACTACCTACCGTGGGAGCTTGCCTCCGAC 314

Db 566 CCCCCCGCGCGCCC 507

Qy 315 ATCTTTCTTCTATCCCAATCGCAGGCGAGGACAGCGGCGCAATCGCGCACTA 374

Db 506 CCCCCCCCCCGGTCGCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 447

Qy 375 TCCTCTCTCTCATAGCGTGTCTCCCGCGCTTCTCAGCGCTCAGTCTCAGCCGAGCC 431

[illegible]

Db 444 CGCGCGSSCCCCCTTCGAGSCCCCTCCGCGSCCTCTCTCCCGCCCGCCCCCCCC 503

Qy 440 CGCGCGCTCCCGGGACCTGTGTGTGACTTCAACAA 476

Db 504 CGGSCCCCCCTTCGCMCCCGCCCGCGSCCTCTCCACCA 540

RESULT 9

EX319543 727 bp mRNA linear EST 07-APR-2003

LOCUS BX319543:AGENAE Rainbow trout multi-tissues subtracted library

DEFINITION (tcay) Oncorhynchus mykiss cDNA clone tcay0037b.k.13 5prim, mRNA sequence.

ACCESSION BX319543

KEYWORDS BX319543.1 GI:29600188

SOURCE EST.

ORGANISM Oncorhynchus mykiss (rainbow trout)

ONCORYNCHUS MYKISS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 727)

REFERENCE Govoroun,M., Guiguen,Y. and Le Gac,F.

AUTHORS Construction and primary characterization of normalized cDNA

TITLE libraries in rainbow trout, Oncorhynchus mykiss

JOURNAL Unpublished

COMMENT Contact: Guiguen Y

INRA - SCRIBE

Campus de beaulieu, RENNES cedex, 35042, France

Tel: 02.23.48.50.09

Fax: 02.23.48.50.20

Email: Yann.Guiguen@beaulieu.rennes.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us at eigenasupport@jouy.inra.fr to obtain the chromatogram of this sequence.

Plate: 0037 row: k column: 13

Seq primer: M13F.

FEATURES

source

1. 727

Location/Qualifiers

/organism="Oncorhynchus mykiss"

/mol_type="mRNA"

/db_xref="taxon:8022"

/clone="tcay0037b.k.13"

/tissue_type="adipose tissue, blood, brain, differentiating gonads, gills, interrenal, intestine, kidney, liver, muscle, ovary, pituitary, testis"

/dev_stage="from embryos to adults"

/lab_host="DH10B"

/clone_lib="AGENAE Rainbow trout multi-tissues subtracted library (tcay)"

/notes="vector: p773D-pac; Clone distribution : AGENAE Resource centre. Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du genome (LRBG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex, FRANCE"

BASE COUNT 206 a 209 c 193 g 118 t 1 others

ORIGIN

Query Match 3.48; Score 52.2; DB 13; Length 727;

Best Local Similarity 55.9; Pred. No. 0.021;

Matches 99; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 715 TTCCCTTGGCGGAACTGTAAACAGACTTATCTTCATGCCAAGCTCTCAAGCGCATCTG 774

Db 471 TACATATGCCCGAGTGTCAAGCTCTCTTCAACCAAGCGACCATCTCAAGTCCACATG 530

Qy 775 CTACGCCACACTGTGTGACCGCCCGTACATGTGTCTTTTCAAGACACCTTCTCTGCG 834

Db 531 CGCTTCCACACGGGTGAGCGCGCTACAAGTGCCAGCAATGTGACAGTCTTCAACCAC 590

Qy 835 AGTGATATCTGAAACGTCTTTTCCAAAAATGCTCAATCAGCGGTGTGAACCCACC 891

Db 591 AACGTGAGTCTGAAAGCCACGCTCCAGCGCTACCCAGCGCTTGCTGCTACGGGACC 647

RESULT 10
LOCUS CC315110
DEFINITION TAM32-30A5 Sp6.1 TAM32 Gallus gallus genomic clone TAM32-30A5, genomic survey sequence.
ACCESSION CC315110
VERSION CC315110.1 GI:30709165
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus
1 (bases 1 to 1137)
Kremitzki,C., Higginsbotham,J., Wylie,K., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
Gallus gallus BAC End Reads
Unpublished
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: Sp6 ATTGAGTGACACTATAG
Class: BAC ends
High quality sequence start: 155
High quality sequence stop: 352.
Location/Qualifiers
1..1137
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-30A5"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="TAM32"
/note="Vector: pECBAC1; Site 1: EcoRI; Site 2: EcoRI; TAM32 Female Chicken library - for library and clone ordering information: http://www.hbz.tamu.edu"
BASE COUNT 189 a 587 c 157 g 178 t 26 others
ORIGIN
Query Match 3.3%; Score 51.4; DB 29; Length 1137;
Best Local Similarity 43.6%; Pred. No. 0.039;
Matches 199; Conservative 0; Mismatches 257; Indels 0; Gaps 0;
Qy 20 CCTCTCGGCTCCGCTTACCAGTCTGCTCAAGGATCGGCGCTCTTCTTCTGTCCTCCA 79
Db 514 CCACGACCCGCTCGCGGCCCCCAGCACCGCGCCCCGACACCTTCCCTCCGCTCCCG 573
Qy 80 TCTGAGATGCGAATGCTCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 139
Db 574 CCCCACCCCCCGCCACACCCCGCCACGACACACACACACACACACACACACACAC 633
Qy 140 ACACCTGCTCTCTCTGCAACCCCATCATAGCAGTCCGCGCTCTCTCTCTCTCTCT 199
Db 634 GCC 693
Qy 200 GGCAGCCGCGCTACCGGCTGATCTGAACAGTACCGCGCATCAAGTACCATGTTTACG 259
Db 694 CGACACCCCG 753
Qy 260 CGTCTTCTGTCGCGCGGATATGCCCCACATACGTTGGGGAGTGTGCTCCGACATTT 319
Db 754 CTGTGTTCTCTCTCCGNNACNNACNNACNNACNNACNNACNNACNNACNNACNN 813
Qy 320 TCCCTTCTCATCCCAATCCGAGGCGGAGGACAGGCGGACCATTCGCGGACATCTCT 379
Db 814 GCCCGCGCC 873

380 CTCCTCATAGCGTCTCCCGCGCTTCCAGCGCTCAGTCTACCGCAGCAATTGGCC 439
Db 874 ACACCCACCCCG 933
Qy 440 CGGCGCTTCCCGGAGCGTGTGCTGCTGACTTCAACA 475
Db 934 GGCACCCCG 969

RESULT 11
LOCUS AZ895651/c
DEFINITION RPCI-24-211J20.TJ RPCI-24 Mus musculus genomic clone RPCI-24-211J20 genomic survey sequence.
ACCESSION AZ895651
VERSION AZ895651.1 GI:13214596
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 567)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsagaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished
Other GSSs: RPCI-24-211J20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 211 row: J column: 20
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..567
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-211J20"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using XhoI partially digested male C57BL/6J DNA."
BASE COUNT 189 a 93 c 254 g 31 t
ORIGIN
Query Match 3.2%; Score 50.6; DB 28; Length 567;
Best Local Similarity 53.2%; Pred. No. 0.047;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
Qy 19 CCTCTCGGCTCCGCTTACAGTCTGCTCAAGGATCGGCGCTCTTCTCTGCTCC 78
Db 343 CCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTCT 284
Qy 79 ATCTCGAGCATGCCATGCTCAGTACAGTACAGTACAGTACAGTACAGTACAGT 138
Db 283 CTGCTCTCTGCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTCT 224

QY 139 CACACCTGCTCTCTGCAACCCCATATAGCAAGTGGCCGCTCTCTACTCTGTAATG 198
 DB 223 CTCTCTGCTCTCTGCTCTCTCTGCTCTCTCTGCTCTCTGCTCTGCTCTGCTATG 164
 QY 199 GGGCAGCGCGCTACCGGCT 219
 DB 163 CTCTGCTCTGCTCTGCT 143

RESULT 12
 BX457568 1201 bp mRNA linear EST 22-MAY-2003
 LOCUS
 DEFINITION BX457568 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
 CS0DA001Y106 3-PRIME, mRNA sequence.
 ACCESSION BX457568
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1201)
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 7353.r for more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DA001BE03NP1&cluster=7353.r. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DA001BE03NP1.

FEATURES
 source
 Location/Qualifiers
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DA001Y106"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."
 BASE COUNT 227 a 498 c 128 g 219 t 129 others
 ORIGIN
 Query Match 3.2%; Score 50; DB 13; Length 1201;
 Best Local Similarity 37.3%; Pred. No. 0.091;
 Matches 179; Conservative 43; Mismatches 258; Indels 0; Gaps 0;
 QY 15 CCATCCCTCTGGGCTCGTCTACAGTCTGCTCAAGAGTGGCGGCTCTCTCTGTCG 74
 DB 244 CSCTTMCAGGCCCCATGCMCCGCCGCCCTCCYAGGCTCACTGCACTCCACCCCYC 303
 QY 75 TCCATCTCGAGATGCGATGCTCTAGTACACATGAGCTCAGTACCACTCTCTCA 134
 DB 304 ACCAAAGCTTACCTTGGCATCTCCCTTTCACGSCCCCTCCMCCCTCCCAAACTCCC 363
 QY 135 GCCGACACCTGCTCTCTGCAACCCCATATAGCAGTGGCGGCTCTCTACTGTA 194
 DB 364 CCCTCATCCCCCGGCTCTCTCTCCCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 423
 QY 195 CATGGGAGCGCGCTGATCGAGGCTGATCTGAACAGGTACCCCGCATCAAGTACAGT 254
 DB 424 AGCCGCGCGGCTCTCAACCCCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 483

QY 255 TTACCGTCTTCTGCTGGCGGATATGCCCACACTACCTGCGGAGCTTCTGCTCCGAC 314
 DB 484 TACMYCCCCCSCCCCCCGCCYCCCGCGGCTCMCCCCCCCCCCCCCCCCCCCC 543
 QY 315 ATCTTTCTTCTCTATCCCAATCCGAGCGGAGGCGAGGCGAGCAATCGCGCACTA 374
 DB 544 AYCC 603
 QY 375 TCCTCTCTCTATAGCGTGTCCCGCGCTTCAGCGTCTAGTCTGCTACCGGCAAT 434
 DB 604 CCTCTCCCGCGGCGSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 663
 QY 435 TCGCGCGGCGCTCCCGGAGCGTCTGCTCACTTCAACAATGAGTCTCTCTCAGGAGC 494
 DB 664 CCGCGGCTCCGCTCCACCASCACAGTATTCACAGTCTAATGCAATCACTGATGCG 723

RESULT 13
 BX421672 1201 bp mRNA linear EST 15-MAY-2003
 LOCUS
 DEFINITION BX421672 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA
 clone CS0DH003YC20 5-PRIME, mRNA sequence.
 ACCESSION BX421672
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1201)
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 1039.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DH003BB10QPI&cluster=1039.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DH003BB10QPI.

FEATURES
 source
 Location/Qualifiers
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DH003YC20"
 /tissue_type="T CELLS (JURKAT CELL LINE)"
 /cell_line="JURKAT CELL LINE"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."
 BASE COUNT 156 a 536 c 203 g 216 t 90 others
 ORIGIN
 Query Match 3.2%; Score 49.6; DB 13; Length 1201;
 Best Local Similarity 38.8%; Pred. No. 0.12;
 Matches 184; Conservative 33; Mismatches 257; Indels 0; Gaps 0;
 QY 14 ACCATCCCTCTGGGCTCGTCTTACAGTCTGCTCAAGGATCGGCGCTCTCTCTCTG 73
 DB 133 MCMCCGCMYYTCMCCCTCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 192
 QY 74 CTCCCATCTCGAGCATGCGCAATGCTCAGTACAGTACAGTACAGTACAGTACAGT 133

[illegible]

RESULT 15
A0745021/c

LOCUS	DEFINITION	HS_5507_Al_H09.SP6.RPCI-11 Human Male BAC Library Homo sapiens genomic clone [Plate=1083 Col=17 Row=O, genomic survey sequence.	AQ745021	786 bp	DNA linear	GSS 16-JUL-1999
NR_022277.1						

ACCESSION AQ745021
VERSION AQ745021.1 GI:5523466
KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 786)

AUTHORS
Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D. and Hood, L.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOHNPAV.	Proc Natl Acad Sci U S A 95 (17) 9739-9744 (1998)

JOURNAL
MEDLINE
PUBMED
COMMENT
FROG. NATL. ACAD. SCI. U.S.A. 96 (1/), 9/39-5/44 (1999)
99380589
10449764
Contact: Mahalingam CV Wallingford Wood 1
Contact: Mahalingam CV Wallingford Wood 1

CONTACT: MANAIRAS GS, WALLACE JC, HOOD L
High Throughput Sequencing Center
University of Washington
401 Ocean View Building North
Seattle, WA 98195

401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

(pietere@jeong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://bacpac.med.buffalo.edu/bac_end_web_server.htm

http://www.htsc.washington.edu
Plate: 1083 row: 0 column: 17

Seq primer: SP6
Class: BAC ends
High quality sequence stop: 786.

100

FEATURES
source Location/Qualifiers
1. .786
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=1083 Col=17 Row=0"
/sex="male"
/clone_lib="RPC1-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI MethyIase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 67 a 42 c 386 g 54 t 237 others
ORIGIN

Query Match 3.2%; Score 49.2; DB 28; Length 786;
Best Local Similarity 28.0%; Pred. No. 0.12;
Matches 165; Conservative 0; Mismatches 425; Indels 0; Gaps 0;

QY 15 CCATCCCTTCGGGCTCGCTACAGTCTGCTCAAGGATCGGCGCTTCTCTCTGC 74
|||
Db 753 CCCCGCCCCCNCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 694
|||
QY 75 TCCCATCTCGAGCATGCCATGCTCAGTACAGATGCGAGCTCAGTACCGAGTCTCTCA 134
|||
Db 693 CCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 634
|||
QY 135 GCGGCACACCTGCTCTCTGCAACCCCATCATAGCAGTCCGCGCTCTCTCACTCGTA 194
|||
Db 633 NNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 574
|||
QY 195 CATGGGCGAGCGCGGTACCGGCTGATCTGAACAGGTACCCCGCATCAAGTCAAGTGT 254
|||
Db 573 NCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 514
|||
QY 255 TTAGGGGTCTTCTGCTGCGCGGATAATGCCCCACACTACCGTGGGAGCTTGCTCCGAC 314
|||
Db 513 NCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 454
|||
QY 315 ATCTTTCTCTTCTCATCCCAATCCGAGGCGAGGCACAGGCGCAGCAATCGCGCACATA 374
|||
Db 453 CCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 394
|||
QY 375 TCTCTCTCTCATAGGTGCTCCGCGCGGCTTCAGGCGTCAAGTACCGGAGCGCAAT 434
|||
Db 393 NCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 334
|||
QY 435 TCGCGCGGCGCTCCCGGGACCGTGTGCTGACTTCAACAATGGACTTCTTCAGGAGC 494
|||
Db 333 CCCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 274
|||
QY 495 ATTCAAGTTATTCGAGCGGAAAGCTCAAGGTTGGGACCCCGTTGCTGCGAATGGTCTGC 554
|||
Db 273 NNN 214
|||
QY 555 GCCGTATCCCGGAAGGAGCTCCCCCGGAAACCCAGGTTGTTGGTCTCAGG 604
|||
Db 213 NNN 164
|||

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 18:04:01 ; Search time 468 Seconds

(without alignments)
8980.823 Million cell updates/sec

Title: US-10-029-180-7

Perfect score: 1557

Sequence: 1 atggatcctagaaccattcc.....cgttcagccgacgacaa 1557

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 13Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1557	100.0	1557	24	ABQ94220 FLO11 gene expres
2	1557	100.0	1800	24	ABQ94276 Fungal gene expres
3	1557	100.0	1833	24	ABQ94275 Fungal gene expres
4	391.6	25.2	867	25	ABZ51240 Aspergillus oryzae
5	280.8	18.0	612	25	ABZ54133 Aspergillus oryzae
6	59	3.8	236	21	NA257129 Pinus radiata tran
7	49.2	3.2	1856	23	ABL28791 Drosophila melanog
8	49.2	3.2	3219	23	ABL28790 Drosophila melanog

C	9	49.2	3.2	6530	14	AAQ51557	Loricrin gene. Ho
C	10	49.2	3.2	6530	20	AAZ22072	Nucleotide sequenc
C	11	49.2	3.2	6530	22	RAC58948	Mouse loricrin gen
	12	48.6	3.1	3073	25	AAC51683	Human nucleic acid
	13	46.6	3.0	549	21	AAC93730	Cat flea hindgut a
	14	46	3.0	4403765	22	AAI99683	Mycobacterium tube
	15	46	3.0	4411529	22	AAI99682	Mycobacterium tube
	16	45.2	2.9	492	17	AAT08579	Partial genomic DN
	17	45.2	2.9	720	22	AAQ50543	Mammalian vestibul
	18	45.2	2.9	1476	17	AAT08578	Zinc finger protei
	19	44.6	2.9	3269	23	ABL26283	Drosophila melanog
	20	44.6	2.9	12920	23	ABL26282	Drosophila melanog
	21	44.4	2.9	4409	23	ABL17895	Drosophila melanog
	22	44.4	2.9	5021	23	ABL11783	Drosophila melanog
	23	44.4	2.9	7394	23	ABL17894	Drosophila melanog
	24	44.4	2.9	7614	23	ABL11782	Drosophila melanog
	25	44.4	2.9	10732	21	AAA10594	Gene encoding a su
	26	44.4	2.9	31814	25	AAZ47150	Human Ras-like pro
	27	43.8	2.8	2480	23	ABL25810	Drosophila melanog
	28	43.8	2.8	2874	23	ABL28677	Drosophila melanog
	29	43.8	2.8	5125	23	ABL28676	Drosophila melanog
	30	43.2	2.8	646	22	AAS26107	Human cDNA encodin
	31	43.2	2.8	646	25	ABX73448	Human novel polynu
	32	43	2.8	336	20	AAV86698	EST clone BG228.
	33	43	2.8	1506	23	ABL29105	Drosophila melanog
	34	43	2.8	2336	21	AAZ35052	Drosophila melanog
	35	43	2.8	2651	21	AAZ35053	Mouse CNREB-2S (sh
	36	43	2.8	3215	23	ABL29108	Mouse CNREB-2L (lo
	37	43	2.8	3506	23	ABL29104	Drosophila melanog
	38	43	2.8	3847	21	AAZ35044	Drosophila melanog
	39	43	2.8	4841	22	AAK52955	Mouse CNREB-2 cDNA
	40	43	2.8	4880	22	AAK51971	Human polynucleoti
	41	43	2.8	4898	25	ABZ79886	Human polynucleoti
	42	43	2.8	16080	24	AAD28651	Human nucleic acid
	43	42.6	2.7	1134	23	ABL25805	Human Sal2 gene.
	44	42.6	2.7	1152	23	ABL13965	Drosophila melanog
	45	42.6	2.7	3213	23	ABL13964	Drosophila melanog

ALIGNMENTS

RESULT 1

ABQ94220

ID ABQ94220 standard; DNA; 1557 BP.

XX ABQ94220;

AC ABQ94220;

DT 22-OCT-2002 (first entry)

DE FLO11 gene expression regulator An09 coding sequence.

XX Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW PC05; PC06; PC07; PC08; PC09; PC10; PC18; PC23; PC24; PC25; PC33; PC34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An000; At167; At221; At233; At239; At240; At274; PC1000; PC1001;
KW loVF; loVE; lovastatin; PC804; acvA; penicillin; antifungal; gene; ds.

OS Unidentified.

XX WO200257456-A2.

PN 25-JUL-2002.

XX 24-DEC-2001; 2001WO-US49911.

XX 22-DEC-2000; 2000US-257431P.

XX (MICR-) MICROBIA.

XX Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;

ID XX ABO94276 standard; DNA; 1800 BP.
AC XX ABO94276;
DT XX 22-OCT-2002 (first entry)
DE XX Fungal gene expression regulator coding sequence SEQ ID 118.
KW XX Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At19; At20; At22; At24; At27; At32;
KW PC05; PC06; PC07; PC08; PC10; PC18; PC23; PC24; PC25; PC33; PC34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; PC1000; PC1001;
KW lovF; lovE; lovastatin; PC804; acvA; penicillin; antifungal; gene; ds.
OS XX Unidentified.
XX WO200257456-A2.
XX 25-JUL-2002.
XX 24-DEC-2001; 2001WO-US49911.
XX 22-DEC-2000; 2000US-257431P.
XX (MICR-) MICROBIA.
XX Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
XX Holtzman D, Sherman A;
XX WPI; 2002-627368/67.
XX New isolated or recombinant gene, or purified protein, useful in
XX regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
XX production of enzymes, secondary metabolites or other commercially and
XX medically useful products
XX Claim 67; Page 66-67; 71pp; English.
XX The present invention relates to novel fungal gene expression regulators
XX (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
XX An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
XX At14, At16, At18, At19, At20, At22, At24, At27, At32, PC05, PC06, PC07,
XX PC08, PC09, PC10, PC18, PC23, PC24, PC25, PC33 and PC34 are FLO11 gene
XX expression regulators. FLO11 is required for fungal invasion and its
XX expression is believed to be regulated by factors that also modulate
XX secondary metabolite production. At279, At286, At291, At320, At322,
XX An1000, At167, At221, At233, At239, At240, At274, PC1000 and PC1001 are
XX lovF gene expression regulators, and At501 and At574 are lovE gene
XX expression regulators. lovF and lovE are believed to be involved in the
XX production of the secondary metabolite lovastatin. PC804 is an acvA gene
XX expression regulator. acvA is involved in the production of the secondary
XX metabolite penicillin. The fungal gene expression regulators and their
XX coding sequences are useful in regulating or manipulating the expression
XX of fungal genes that are involved in the production of enzymes, secondary
XX metabolites and other commercially and medically useful products, in
XX order to achieve maximum benefit. The genes may also be used to identify
XX genes relevant to fungal invasion which may act as targets for the
XX development of antifungal drugs.
XX Sequence 1800 BP; 372 A; 606 C; 463 G; 359 T; 0 other;
XX Query Match 100.0%; Score 1557; DB 24; Length 1800;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGATCCTAGAAACCATCCCTCTCGGCTCCGCTACCACTGCTCGCTCAAGGATCGGCG 60
DB 244 ATGGATCCTAGAAACCATCCCTCTCGGCTCCGCTACCACTGCTCGCTCAAGGATCGGCG 303
QY 61 CCTCTTCTTCTGTCTCCCACTCGAGATGCGCAATGCTCAGTACAGATGCGCTCAG 120

DB 304 CCTCTTCTTCTGTCTCCCATCTCGAGCATGCAATGCTCAGTACAGATGCGCTCAG 363
QY 121 TACCAGTCTCTCAGCGCGCACACCTGCTCTCTCTGCAACCCCATCATAGCAGTGGGCC 180
DB 364 TACCAGTCTCTCAGCGCGCACACCTGCTCTCTCTGCAACCCCATCATAGCAGTGGGCC 423
QY 181 GTCCTCAGTCTGATGCGGCGCGCGCTGACCGGCTGATCTGAAACAGGTACCCGCA 240
DB 424 GTCCTCAGTCTGATGCGGCGCGCGCTGACCGGCTGATCTGAAACAGGTACCCGCA 483
QY 241 TCAAGTCAAGATGTTTACGGGTCTTCTGCTGCGCGCGCAATATGCCCCACACTTACCGTGGC 300
DB 484 TCAAGTCAAGATGTTTACGGGTCTTCTGCTGCGCGCGCAATATGCCCCACACTTACCGTGGC 543
QY 301 AGCTTGCTCGGACATCTTCTTCTCTATCCCAATCCGAGGCGGACAGGCGGAG 360
DB 544 AGCTTGCTCGGACATCTTCTTCTCTATCCCAATCCGAGGCGGACAGGCGGAG 603
QY 361 CAATCGCGGCACTATCTCTCTCTATGAGGTGCTCCCGCGGCTTCCAGGCTCAGTGC 420
DB 604 CAATCGCGGCACTATCTCTCTCTATGAGGTGCTCCCGCGGCTTCCAGGCTCAGTGC 663
QY 421 TACCGGAGCAATGCGCGGCGCTTCCCGGAGCGGTCTGCTGCTGATCTTCAACAATGGA 480
DB 664 TACCGGAGCAATGCGCGGCGCTTCCCGGAGCGGTCTGCTGCTGATCTTCAACAATGGA 723
QY 481 CTTCCTTCAAGAGATTCAGTTATTTGGAGCGGAAAGCTCAAGTTGGGACCCCGTTCCT 540
DB 724 CTTCCTTCAAGAGATTCAGTTATTTGGAGCGGAAAGCTTCAAGTTGGGACCCCGTTCCT 783
QY 541 GCGAATGCTGCTCGGCGGTATCCCGGAGAGACTCCCGCGGAAAGCTTCCCGCGGAAAGCTTGGTGTCT 600
DB 784 GCGAATGCTGCTCGGCGGTATCCCGGAGAGACTTCCCGCGGAAAGCTTCCCGCGGAAAGCTTGGTGTCT 843
QY 601 CAGGGGCGAGCGGATCTCTTCCGAGTGTTCGGGAGCGCAACTCCGGTCAACAATGCT 660
DB 844 CAGGGGCGAGCGGATCTCTTCCGAGTGTTCGGGAGCGCAACTCCGGTCAACAATGCT 903
QY 661 GTTAAAGGCGAGCGGAGAGACTTATCCCGCGGCAAGATGCGGATGGAAGTTCCT 720
DB 904 GTTAAAGGCGAGCGGAGAGACTTATCCCGCGGCAAGATGCGGATGGAAGTTCCT 963
QY 721 TGCCCGAAGCTGTAAAGAGCTTATCTTCAAGCAAGCATCTCAAGCGCCATCTGTACGC 780
DB 964 TGCCCGAAGCTGTAAAGAGCTTATCTTCAAGCAAGCATCTCAAGCGCCATCTGTACGC 1023
QY 781 CACACTGTGAGCGCGCTGATGCTGTGTTCTTTGCAAGAGACACTTCTCTCGAGTGT 840
DB 1024 CACACTGTGAGCGCGCTGATGCTGTGTTCTTTGCAAGAGACACTTCTCTCGAGTGT 1083
QY 841 ATCTGAAAGCTCATTTCCAAAATGCTCAATCAGGCGGTGTAACCCCGGAGCAAG 900
DB 1084 ATCTGAAAGCTCATTTCCAAAATGCTCAATCAGGCGGTGTAACCCCGGAGCAAG 1143
QY 901 CACTGTGCGCACCCCAATGCGATGTAAGAGGTCCCAACAGCAGGCTGCGGCGAATCT 960
DB 1144 CACTGTGCGCACCCCAATGCGATGTAAGAGGTCCCAACAGCAGGCTGCGGCGAATCT 1203
QY 961 GTAAAACTGTCCAGGATGAAGTCAAGTGTGCTTCCCGCTCCCAATGCGATCCCGGGC 1020
DB 1204 GTAAAACTGTCCAGGATGAAGTCAAGTGTGCTTCCCGCTCCCAATGCGATCCCGGGC 1263
QY 1021 AGCACTTACGGGAGGAGCGCTCAACGGCAATGGAATGCTCCGGCGCGGCGAGGTAC 1080
DB 1264 AGCACTTACGGGAGGAGCGCTCAACGGCAATGGAATGCTCCGGCGCGGCGAGGTAC 1323
QY 1081 GCGGATCACAGACTATGGCTTCCCAATGCTATCGTCAAGCGGATGCGCGGTGTGTCAG 1140
DB 1324 GCGGATCACAGACTATGGCTTCCCAATGCTATCGTCAACGGGATGCGCGGTGTGTCAG 1383
QY 1141 CTTGAAAGACGCTTTCCCGGCGGCGCGGCGCATCAAGAGGCGGCTTTGGCCACAGCTCCC 1200
DB 1384 CTTGAAAGACGCTTTCCCGGCGGCGGCGGCGCATCAAGAGGCGGCTTTGGCCACAGCTCCC 1443

QY	1201	AAGCAGAGCCCGGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAAGTTGAATATT	1260
DB	1444	AAGCAGAGCCCGGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAAGTTGAATATT	1503
QY	1261	GACCGAAACATCAGAGCAGGTAAACAACCCGGTTGTTCAAGACCCCAAGCCCTGTGTATG	1320
DB	1504	GACCGAAACATCAGAGCAGGTAAACAACCCGGTTGTTCAAGACCCCAAGCCCTGTGTATG	1563
QY	1321	CCAGGACATCCCGGCCACCCCGGTGACTTGAGCGTCTATGTTCCAACTCAAGCT	1380
DB	1564	CCAGGACATCCCGGCCACCCCGGTGACTTGAGCGTCTATGTTCCAACTCAAGCT	1623
QY	1381	CCCGAGGGTACATGTTCTCCAGTCTATGCTTGGTGGTCAAGAGCCCATCCACGCTCAT	1440
DB	1624	CCCGAGGGTACATGTTCTCCAGTCTATGCTTGGTGGTCAAGAGCCCATCCACGCTCAT	1683
QY	1441	GTCCGAGACCCGAGCGAAGTATTACCCACCACTACCGCTGTCAGAGAGTGGAATGAAC	1500
DB	1684	GTCCGAGACCCGAGCGAAGTATTACCCACCACTACCGCTGTCAGAGAGTGGAATGAAC	1743
QY	1501	GGTCTCTATCTGGCTTCGACTATGAGTGGGACGCGCACCGTTCAGCCCGCCAGACAA	1557
DB	1744	GGTCTCTATCTGGCTTCGACTATGAGTGGGACGCGCACCGTTCAGCCCGCCAGACAA	1800
RESULT 3			
ABQ94275			
ID	ABQ94275	standard; DNA; 1833 BP.	
XX	ABQ94275;		
AC	ABQ94275;		
XX			
XX			
DT	22-OCT-2002	(first entry)	
XX			
DE	Fungal gene expression regulator coding sequence	SEQ ID 117.	
XX			
KW	Fungal gene expression regulator; fungicide; gene therapy; An01; An05;		
KW	An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;		
KW	At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;		
KW	Pc05; Pc06; Pc07; Pc08; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;		
KW	FL011; fungal invasion; secondary metabolite; At279; At286; At291; At320;		
KW	At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;		
KW	lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.		
XX			
OS	Unidentified.		
XX			
XX	WO200257456-A2.		
PN			
XX			
PD	25-JUL-2002.		
XX			
PF	24-DEC-2001; 2001WO-US49911.		
XX			
XX	22-DEC-2000; 2000US-257431P.		
PR			
XX			
XX	(MICR-) MICROBIA.		
PA			
XX			
PI	Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;		
FI	Holtzman D, Sherman A;		
XX			
DR	WPI; 2002-627368/67.		

New isolated or recombinant gene, or purified protein, useful in
 regulating fungal gene expression of FLO11, lovP, lovR or acvA for the
 production of enzymes, secondary metabolites or other commercially and
 medically useful products -
 Claim 67; Page 66; 71pp; English.

PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX WPI; 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 XX
 PS Claim 1; Page 592; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/WADS, homeodomain zipper, LIM domain, AP2
 CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 236 BP; 57 A; 82 C; 61 G; 36 T; 0 other;
 Query Match 3.8%; Score 59; DB 21; Length 236;
 Best Local Similarity 61.3%; Pred. No. 6.2e-07;
 Matches 95; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 709 GGAAGTTCCTTGGCCGCACTGTAAAGACTTATCTTCATGCCAAGCATCTCAAGCGC 768
 Db 7 GGAGACCAACGCTTGCCCGCTGTGCCAAGCGACTTTTACTCGCCGCAACATGTCGACGA 66
 QY 769 CATCTGCTACGCACACACTGGTGAACCGCCGTCATGTGTCTTTCGAAGACACCTTC 828
 Db 67 CACATGCGCTCCACACCGGCGACCGCCGTCACAAAGTGTCTCCATCTGCACCGACTCGTT 126
 QY 829 TCTCCAGTGTATCTCTGAAAGCTCATTTCCAAAA 863
 Db 127 GGCCGACGCGACTCTCTGAGGCGCATGAGAGAA 161
 RESULT 7
 ABL28791
 ID ABL28791 standard; DNA; 1856 BP.
 AC ABL28791;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37846.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 37846; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABE57737-ABE72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1856 BP; 532 A; 451 C; 449 G; 424 T; 0 other;
 Query Match 3.2%; Score 49.2; DB 23; Length 1856;
 Best Local Similarity 59.2%; Pred. No. 0.0011;
 Matches 84; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
 QY 714 GTTCCCTTGGCCGCACTGTAAAGACTTATCTTCATGCCAAGCATCTCAAGCGCATCT 773
 Db 1307 GTATCTCTGCAAAATGTGCAGCAAGAGCTTTGTGAGTCCAAACGACCTTAAGATCCATAT 1266
 QY 774 GCTACGGCACACTGGTGACCGCCGTCATGTGTCTTTTGAAGACACCTTCTCTCG 833
 Db 1367 GCGCGCCACACACGCGGAGGACCATATCAGTGTGGAGTGTGCGCGAGAGCTTTGTGTG 1326
 QY 834 CAGTGATATCTGAAAGCTCAT 855
 Db 1327 CGGTTCGACCTGAATATCCAT 1348
 RESULT 8
 ABL28790/C
 ID ABL28790 standard; DNA; 3219 BP.
 XX
 AC ABL28790;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37843.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 37843; 21pp + Sequence Listing; English.
 XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3219 BP; 842 A; 747 C; 731 G; 899 T; 0 other;

Query Match 3.2%; Score 49.2; DB 23; Length 3219;
Best Local Similarity 59.2%; Pred. No. 0.0014;
Matches 84; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 714 GTTCCCTGCGGCACTGTAAAGACTTATCTTCATGCCAGCATCTCAAGCGCCATCT 773
Db 960 GTACTCTGCAAAATGTGAGGAGAGCTTTGTGAGTCCACGACTTAAGATCCATAT 901
QY 774 GCTACGCCACATGTGTGACCGCCGTACATGTGTGTCTTTGCAAGACACCTTCTCTCG 833
Db 900 GCGGCGCCACACGCGGCGAGCCATATCAGTGTGGAGTGTGCGGAGAGCTTTGTGTG 841
QY 834 CAGTGATATCTGAAAGTCAT 855
Db 840 CGGTTGCGACCTGATATCCAT 819

RESULT 9

AAQ51557/c
ID AAQ51557 standard; DNA; 6530 BP.

XX AC AAQ51557;

XX 25-MAR-2003 (updated)
DT 17-MAY-1994 (first entry)

XX DE Loricrin gene.

XX Loricrin; gene; constitutive; inducible; vector; 5' flanking region;
XX intron; 3' flanking region; TATA box; cap site; intron/exon boundary;
XX DNA cassette; linker; restriction site; induction; human; psoriasis;
XX epidermal cell; wound healing; ulcers; growth factor; antisense RNA;
XX transforming growth factor beta; cytokine; skin; tracheal epithelia;
XX cancer; E6; E7; papilloma virus; p53 protein; vaccine; oral epithelia;
XX capsid protein; squamous epithelium; viral; oesophageal epithelia;
XX vaginal epithelia; corneal epithelia; transduction; transplant; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
FT misc_RNA 1..1540

FT /tag= a
FT /note= "5' flanking region"
FT intron 1587..2677

FT /tag= b
FT /note= "Intron and intron/exon boundary"
FT misc_RNA 4384..6530

FT /tag= c
FT /note= "3' flanking region"

XX W09322431-A1.

XX 11-NOV-1993.

XX 28-APR-1993; 93WO-US03993.

XX 30-APR-1992; 92US-0876286.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX

FI Greenhalgh DA, Roop DR, Rothnagel JA;

XX WPI; 1993-368790/46.

XX Constitutive and inducible vectors based on loricrin or keratin
PT K6 genes - for selective expression in the epidermis, e.g. for
PT wound healing, treating psoriasis and skin cancer, etc.

XX Claim 23; Page 26-32; 68pp; English.

XX This sequence represents the loricrin gene and was used in the
CC construction of the loricrin constitutive vector of the invention.
CC The loricrin vector has a 5' flanking region of approx. 1.5 kb, an
CC intron of approx. 1.1 kb and a 3' flanking region of approx. 1.5 kb.
CC The 5' flanking region of the loricrin gene comprises a TATA box,
CC cap site, first intron and intron/exon boundary, all in the correct
CC order and position for expression of a nucleic acid cassette. The
CC loricrin vector also comprises a 3' flanking region of the loricrin
CC gene and a linker with a unique restriction site at the position of
CC the start and stop codons, joining the two flanking sequences and
CC also having a position for insertion of the DNA cassette (see also
CC AAQ51558). Vectors such as this can be used for; (a) in vitro induction
CC of human epidermal cells, especially for stimulating healing of
CC wounds, surgical incisions and ulcers, (where the vector expresses a
CC growth factor), (b) treating psoriasis, (the vector expressing
CC transforming growth factor beta, a soluble cytokine or antisense
CC RNA), (c) treating skin cancer, (the vector expressing antisense RNA
CC of the E6 or E7 gene of human papilloma virus or normal p53 protein),
CC (d) for vaccination (the vector expresses a viral capsid protein,
CC especially of human papilloma virus), or (e) treating cancer of the
CC squamous epithelium (of epidermal, oral, oesophageal, vaginal,
CC tracheal or corneal epithelia, the vector expresses antisense RNA).
CC Alternatively human epithelial cells are transduced ex vivo, then
CC transplanted.
XX (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 other;

Query Match 3.2%; Score 49.2; DB 14; Length 6530;
Best Local Similarity 46.1%; Pred. No. 0.002;
Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 96 GCCTCAGTACAGTACGAGCCCTCAGTACCCAGTCTCTCAGCGGCACACCTGCTCTCT 155
Db 3530 GCCTCCGTAGTCTGCGCACTGATCTGTGGACAGATCCGCCGCCGCCGAGGA 3471

QY 156 GCAACCCCATCATAGCCAGTCCCGCTCCTCCTCATCTGATGGGCGAGCCGCTACCG 215
Db 3470 ACCACCTCGCAGTCTGTGCGCGCCGCCACCGAATAGCGCGCGCGCGCAGTGA 3411

QY 216 GCCTGATCTGAACAGGTACCCCGCATCAAGTCAAGTCAAGTCTTGTGTCGCGCC 275
Db 3410 TCCACCGCCACCTGAGTAGCGCGCCGCCACAGTAGAGCGCGCTTCCACCGCGCC 3351

QY 276 GATATGCCCCACACTACCGTGGGCGAGCTTGCTCCGACATCTTCTCTCATCCCAA 335
Db 3350 GGAATAGCGCGCTCCGCGAGCTGGAGCCACCGCGGCCACACAGAGTACTTGCCTCCGGA 3291

QY 336 TCCGCGAGCGGAGGACACAGGGCGAGCAATCGCGCATCTATCTCTCTCATAGGTGCT 395
Db 3290 ACCGCGCGCGAGCCACCGGAGCTGCGGCTCCAGAGGAGCTCCGCGAGTAGGAGCGCC 3231

QY 396 CCGGCGCGGCTTCCAGCGCTCAGTGTATCCCGAGCAATTTGCGCGCGGCTTCCCGGG 453
Db 3230 GCCTCTGATCGGCTTCGCGAGCTGGAGCGCGCGCGCGCGCGCGCGAGCGCGCG 3173

RESULT 10

AAZ22072/c

ID AAZ22072 standard; DNA; 6530 BP.

XX

AC AAZ22072;

```
XX 25-NOV-1999 (first entry)
XX Nucleotide sequence of the lorcrin gene.
XX Vector; epidermal cell; lorcrin gene; K6 keratin gene; wound;
XX surgical incision; disease; psoriasis; cancer; alopecia;
XX transgenic animal; ss.
XX Mus sp.
XX US958764-A.
XX 28-SEP-1999.
XX 01-NOV-1993; 93US-0146930.
XX 30-APR-1992; 92US-0876286.
XX 29-OCT-1993; 93US-0145388.
XX (BAYU) BAYLOR COLLEGE MEDICINE.
XX Greenhalgh DA, Rothnagel JA, Roop DR;
XX WPI; 1999-561044/47.
XX Gene therapy vectors for expression in the epidermis -
XX Example 1; Columns 25-32; 34pp; English.
XX The specification describes a vector for the expression of a nucleic
XX acid sequence in an epidermal cell, where the nucleic acid sequence is
XX not a mammalian lorcrin gene or a mammalian K6 keratin gene. The vector
XX comprises a 5' flanking region including the sequences required for its
XX expression; a 3' region from a mammalian lorcrin or a mammalian K6
XX keratin gene which regulates expression (predominantly in an epidermal
XX cell) of the nucleic acid sequence with which it is associated; and
XX a linker connecting the 5' flanking region to the nucleic acid
XX sequence, where the linker comprises a position for inserting the
XX nucleic acid sequence and lacks the coding sequence of a gene with
XX which it is naturally associated. The vector is able to encode a hormone,
XX a growth factor, an enzyme, a clotting factor, an apolipoprotein, a
XX receptor, a drug or an antigen. The vectors can be used to treat wounds
XX or surgical incisions as well as diseases, such as psoriasis, cancer,
XX alopecia and to create transgenic animals for assessing human disease
XX in an animal model. The present sequence represents the lorcrin gene.
XX
XX SQ Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 other;
Query Match 3.2%; Score 49.2; DB 20; Length 6530;
Best Local Similarity 46.1%; Pred. No. 0.002;
Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
QY 96 GCCTCAGTACAGATGCGCTCAGTACCCAGTCTCTCAGCGGCACACCTGCTCTCTCT 155
DB 3530 GCCTCCGTAGTCTGTGGCACTGATCTGTGGAGCATCCGCGCGCGCCACCGGAGGA 3471
QY 156 GCACCCCATCATAGCCAGTCCGCGCTCTCTCTCAGCGGCACACCTGCTCTCTCT 215
DB 3470 ACCACTCCGAGCTGTGCGCGCGCCACCGAGATAGCGCGCGCGCGAGCTGGA 3411
QY 216 GCCTGATCTGAACAGGTACCCCGCATCAAGTCAAGTCTTACCGCTTCTGCTGCGCC 275
DB 3410 TCCACCGCCACCTGAGTAGCGCGCCACAGTAGAGCGCGCGCTTCCACCGCGCC 3351
QY 276 GATATGCCCCACATACCTGCTGCGAGCTTGGCTCCGACATCTTCTTCTCATCCCAA 335
DB 3350 GGAATAGCGCGCTCCGAGCTGTGAGCCACCGCGCCACAGAGTACTTTCGCGCTCCGA 3291
QY 336 TCCGAGCGCGGACGACAGCGCGAGCAATCGCGCATCTCTCTCTCTATAGCGTCT 395
DB 3290 ACCGCCCGCGAGCCACCGAGCTGCGCTTCCAGAGAGCGCTCCGAGTAGGAGCGCC 3231
QY 396 CCAGCGCGCTTCCAGCGCTCAGTCTGATCCGCGAGCAATTCGCGCGCGCTCCCGG 453
DB 3230 GCCTCTGATCCGCTCCGAGCTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3173
RESULT 11
AAC68948/c
ID AAC68948 standard; DNA; 6530 BP.
XX AAC68948;
XX 26-FEB-2001 (first entry)
XX Mouse lorcrin gene.
XX Lorcrin; cytostatic; gene therapy; cell proliferation inhibition;
XX murine; skin cancer; skin ulcer; psoriasis; alopecia; ds.
XX Mus sp.
XX US6143727-A.
XX 07-NOV-2000.
XX 05-JUN-1995; 95US-0458240.
XX 01-NOV-1993; 93US-0146930.
XX 30-APR-1992; 92US-0876286.
XX 29-OCT-1993; 93US-0145388.
XX (BAYU) BAYLOR COLLEGE MEDICINE.
XX Roop DR, Greenhalgh DA, Rothnagel JA;
XX WPI; 2001-006441/01.
XX Gene therapy of skin cancer involving administering specific expression
XX vector, lorcrin or K6 keratin expression vector within the epidermis
XX or epidermal cells.
XX Claim 8; Columns 25-32; 35pp; English.
XX The present invention relates to a method for treating skin cancer. The
XX method comprises administering a lorcrin or K6 keratin expression vector
XX comprising a nucleic acid sequence encoding a wild-type p53 protein at or
XX directly around the site of skin cancer cell, where the expression of the
XX p53 protein by the skin cancer cell results in inhibition of the cell's
XX proliferation. The present sequence is the murine lorcrin gene. This
XX gene was used to construct the lorcrin expression vector of the present
XX invention. The method of the present invention may also be used to treat
XX skin ulcers, psoriasis and alopecia.
XX
XX SQ Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 other;
Query Match 3.2%; Score 49.2; DB 22; Length 6530;
Best Local Similarity 46.1%; Pred. No. 0.002;
Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
QY 96 GCCTCAGTACAGATGCGCTCAGTACCCAGTCTCTCAGCGGCACACCTGCTCTCTCT 155
DB 3530 GCCTCCGTAGTCTGTGGCACTGATCTGTGGAGCATCCGCGCGCGCCACCGGAGGA 3471
QY 156 GCACCCCATCATAGCCAGTCCGCGCTCTCTCTCAGCGGCACACCTGCTCTCTCT 215
DB 3470 ACCACTCCGAGCTGTGCGCGCGCCACCGAGATAGCGCGCGCGCGAGCTGGA 3411
QY 216 GCCTGATCTGAACAGGTACCCCGCATCAAGTCAAGTCTTACCGCTTCTGCTGCGCC 275
DB 3410 TCCACCGCCACCTGAGTAGCGCGCCACAGTAGAGCGCGCGCTTCCACCGCGCC 3351
QY 276 GATATGCCCCACATACCTGCTGCGAGCTTGGCTCCGACATCTTCTTCTCATCCCAA 335
DB 3350 GGAATAGCGCGCTCCGAGCTGTGAGCCACCGCGCCACAGAGTACTTTCGCGCTCCGA 3291
QY 336 TCCGAGCGCGGACGACAGCGCGAGCAATCGCGCATCTCTCTCTCTATAGCGTCT 395
DB 3290 ACCGCCCGCGAGCCACCGAGCTGCGCTTCCAGAGAGCGCTCCGAGTAGGAGCGCC 3231
```


XX (GENO-) INST GENOMIC RES.
PA Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ
XX
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=629432881.
XX
XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
SQ
Query Match 3.0%; Score 46; DB 22; Length 4411529;
Best Local Similarity 45.8%; Pred. No. 0.31;
Matches 197; Conservative 0; Mismatches 230; Indels 3; Gaps 1;
QY 27 GCCTCGCTTACAGTCTGCTCAAGGATCGGCGCTCTTCTTGTGCTCCCATCTCTCGAG 86
Db 1631626 GCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1631685
QY 87 CATGCCAATGCCCTCAGTACAG---ATGCAGCTCAGTACCGAGTCTCTCAGCGGCACAC 143
Db 1631686 GCGCGCGTGTGGCGCGCTACCGGAGCAGCGCGCTGCGCGCGCGCGCGCGCGCG 1631745
QY 144 CTGCGCTCTCTGCAACCCCATATAGCAGTCCGCGCTCTCTCACTCGTACATGGGCA 203
Db 1631746 CGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1631805
QY 204 GCGCGCTACCGCGCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTTTACCGCTC 263
Db 1631806 GCGCGCTGCG 1631865
QY 264 TTCTGCTGCGCGGATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323
Db 1631866 ACCG 1631925
QY 324 TTCTCATCCCAATCCG 383
Db 1631926 ACCG 1631985
QY 384 TCATAGCGTCTCCCG 443
Db 1631986 GTTTGGCG 1632045
QY 444 GCCTCCCGCG 453
Db 1632046 GCGCGCGCGCG 1632055

Search completed: January 2, 2004, 19:29:18
Job time : 490 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 18:05:06 ; Search time 5899 Seconds
(without alignments)
10797.811 Million cell updates/sec

Title: US-10-029-180-7
Perfect score: 1557
Sequence: 1 atggatcctagaaccatcc.....ccgttcagccgcagacaa 1557

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vi.*

29: em_vt.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1557	100.0	1557	6	AX505141	AX505141 Sequence
2	1557	100.0	1560	8	AF467711	AF467711 Emericell
3	1557	100.0	1800	6	AX505252	AX505252 Sequence
4	1557	100.0	1833	6	AX505251	AX505251 Sequence
5	236.2	15.2	741	11	PM68112G	AL686183 Penicilli
6	61	3.9	7218	6	I66494	I66494 Sequence 14
C 7	59.6	3.8	241859	2	AC136406	AC136406 Rattus no
C 8	57.2	3.7	214690	2	AC083889	AC083889 Mus muscu
9	57.2	3.7	219559	10	AC087329	AC087329 Mus Muscu
10	56.8	3.6	125020	9	AF429315	AF429315 Homo sapi
11	52.2	3.4	227533	2	AC094895	AC094895 Rattus no
12	52	3.3	178757	10	AC122509	AC122509 Mus muscu
13	51.8	3.3	197714	2	AC128967	AC128967 Rattus no
C 14	51.6	3.3	264103	2	AC126651	AC126651 Rattus no
C 15	51	3.3	5452	6	AX646997	AX646997 Sequence
C 16	51	3.3	298166	2	AC087563	AC087563 Homo sapi
17	50.8	3.3	236054	2	AC114393	AC114393 Rattus no
C 18	50	3.2	210582	2	BC324212	BC324212 Danio rer
19	50	3.2	232885	2	AC106163	AC106163 Rattus no
20	49.6	3.2	87120	2	AC012225	AC012225 Homo sapi
21	49.4	3.2	33058	2	AC100211	AC100211 Mus muscu
C 22	49.2	3.2	694	3	DMHSP70AB	Y14393 Drosophila
C 23	49.2	3.2	1751	10	MUSLRUNA	M34398 Mouse loric
C 24	49.2	3.2	1804	10	BC026781	BC026781 Mus muscu
25	49.2	3.2	1863	3	AY052029	AY052029 Drosophil
C 26	49.2	3.2	6530	6	AR076241	AR076241 Sequence
C 27	49.2	3.2	6530	10	MM009189	U09189 Mus musculu
C 28	49.2	3.2	55528	3	AC105764	AC105764 Caenorhab
C 29	49.2	3.2	144152	3	AC007725	AC007725 Drosophil
C 30	49.2	3.2	182183	3	AC007889	AC007889 Drosophil
C 31	49.2	3.2	226782	3	AE003693	AE003693 Drosophil
C 32	49.2	3.2	255454	2	AC138676	AC138676 Mus muscu
C 33	48.8	3.1	209887	2	AC079424	AC079424 Mus muscu
C 34	48.6	3.1	156506	2	AC068038	AC068038 Homo sapi
C 35	48.6	3.1	194864	10	AL691489	AL691489 Mouse DNA
C 36	48.6	3.1	223879	9	AC008735	AC008735 Homo sapi
37	48.4	3.1	183991	2	BX072578	BX072578 Danio rer
C 38	48	3.1	5880	14	HS84	DL486 Equine herp
C 39	47.4	3.0	188974	2	AC022682	AC022682 Mus muscu
C 40	47.4	3.0	303091	2	AC084799	AC084799 Mus muscu
C 41	47	3.0	161002	2	BX005475	BX005475 Danio rer
42	46.8	3.0	1349	3	AY122249	AY122249 Drosophil
C 43	46.8	3.0	85968	2	AC009536	AC009536 Drosophil
44	46.8	3.0	181773	2	AC136125	AC136125 Rattus no
C 45	46.8	3.0	212493	2	AC080167	AC080167 Mus muscu

ALIGNMENTS

RESULT 1
AX505141
LOCUS AX505141
DEFINITION Sequence 7 from Patent WO02057456.
ACCESSION AX505141
VERSION AX505141.1 GI:23386445
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Cali,B.M., Madden,K.T., Milne,T.G., Zhang,L., Silva,J.C.,
Trueheart,J., Holtzman,D. and Sherman,A.
TITLE Regulators of fungal gene expression
JOURNAL Patent: WO 02057456-A 7 25-JUL-2002;

Microbia (US) Location/Qualifiers
1. .1557
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="fungal gene"
BASE COUNT 329 a 533 c 385 g 310 t
ORIGIN

Query Match 100.0%; Score 1557; DB 6; Length 1557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCTTAGAAACCATCCCTCTCGGCTCCCTCTACGATCTGCTCAAGATCGGG 60
DB 1 ATGGATCTTAGAAACCATCCCTCTCGGCTCCCTCTACGATCTGCTCAAGATCGGG 60

QY 61 CCTCTTCTCTCTCTCCATCTCGAGATGCCAATGCTCAGTACAGATGAGCTCTAG 120
DB 61 CCTCTTCTCTCTCTCCATCTCGAGATGCCAATGCTCAGTACAGATGAGCTCTAG 120

QY 121 TACCCAGTCTCTCAGCGCGCACACCTGCTCTCTGCAACCCCATCATAGCCAGTCCGCC 180
DB 121 TACCCAGTCTCTCAGCGCGCACACCTGCTCTCTGCAACCCCATCATAGCCAGTCCGCC 180

QY 181 GCTCTCTCAGTACATGGGGGAGCGCGCTGATCGGCTGATCTGAAAGATACCGGCA 240
DB 181 GCTCTCTCAGTACATGGGGGAGCGCGCTGATCGGCTGATCTGAAAGATACCGGCA 240

QY 241 TCAAGTACAGATGTTTACGGGTCTTCTGCTGCGCGGATATGCCACATCACTACCGTGGG 300
DB 241 TCAAGTACAGATGTTTACGGGTCTTCTGCTGCGCGGATATGCCACATCACTACCGTGGG 300

QY 301 AGCTTGCTCCGACATCTTCTCTTCTATCCCAATCCGAGCGCGCACAGGCGCGAG 360
DB 301 AGCTTGCTCCGACATCTTCTCTTCTATCCCAATCCGAGCGCGCACAGGCGCGAG 360

QY 361 CAATGCGCGCATATCTCTCTCTCAGGCTGCTCCGCGCGCTTCCAGGCTCAGTCG 420
DB 361 CAATGCGCGCATATCTCTCTCTCAGGCTGCTCCGCGCGCTTCCAGGCTCAGTCG 420

QY 421 TACCGGAGCAATGTCGCGCGGCTTCTGCTGCGGAGCGCTGCTGACTTCAACAATGGA 480
DB 421 TACCGGAGCAATGTCGCGCGGCTTCTGCTGCGGAGCGCTGCTGACTTCAACAATGGA 480

QY 481 CTTCCTTTCAGGAGCATTCAGTTATTCGAGCGGAAGCTCTCAAGTTGGGACCCCGTTGCT 540
DB 481 CTTCCTTTCAGGAGCATTCAGTTATTCGAGCGGAAGCTCTCAAGTTGGGACCCCGTTGCT 540

QY 541 GCGATGCTGTCGCGGCTATCCCGGAGAGGACTCCCGCGGAGCGGCTGCTGTTCT 600
DB 541 GCGATGCTGTCGCGGCTATCCCGGAGAGGACTCCCGCGGAGCGGCTGCTGTTCT 600

QY 601 CAGGGGAGCGGCTATCTTCCAGTGTTCGCGGAGCGCACTCCCGGTCAACAATGTT 660
DB 601 CAGGGGAGCGGCTATCTTCCAGTGTTCGCGGAGCGCACTCCCGGTCAACAATGTT 660

QY 661 GTTAAACGCGGCAAGAACCACTACTATCCCGGCAAGGATCCGATGAAAGTTCCCT 720
DB 661 GTTAAACGCGGCAAGAACCACTACTATCCCGGCAAGGATCCGATGAAAGTTCCCT 720

QY 721 TGCCGGAAGTATCAAGATCTATCTTATGCGCAAGCATCTCAAGCGCATCTGCTACGC 780
DB 721 TGCCGGAAGTATCAAGATCTATCTTATGCGCAAGCATCTCAAGCGCATCTGCTACGC 780

QY 781 CACACTGTGACCGCGCTATCTGCTGCTTCTTTCGAAAGACACTTCTCTCGCAGTAT 840
DB 781 CACACTGTGACCGCGCTATCTGCTGCTTCTTTCGAAAGACACTTCTCTCGCAGTAT 840

QY 841 ATCTGGAACGCTATTTTCCAAAAATGCTCAATCAGGCGTGTGTAACCCACCGGAGCAACG 900
DB 841 ATCTGGAACGCTATTTTCCAAAAATGCTCAATCAGGCGTGTGTAACCCACCGGAGCAACG 900

QY 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGTCCCAACAGCAGCGCTGCGCGAATCT 960
DB 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGTCCCAACAGCAGCGCTGCGCGAATCT 960

QY 961 GTAAAACTGTCCAGGATGAAGTACAGTACCGTCCCGCTTCCCAATGGCATCCCGGC 1020
DB 961 GTAAAACTGTCCAGGATGAAGTACAGTACCGTCCCGCTTCCCAATGGCATCCCGGC 1020

QY 1021 AGGACTTACGGGAGGAGCGCTCAACGGCATGAGTCTAGTCTCCGCGCGGCGGAGGTAC 1080
DB 1021 AGGACTTACGGGAGGAGCGCTCAACGGCATGAGTCTAGTCTCCGCGCGGCGGAGGTAC 1080

QY 1081 GGGGATCACAGACTATGCTTCCCAATGTCCGTCACCGGATGGCGCTGTGTCTAG 1140
DB 1081 GGGGATCACAGACTATGCTTCCCAATGTCCGTCACCGGATGGCGCTGTGTCTAG 1140

QY 1141 CCTGAAGACGGCTTTCGCGGCGCGCGCATCAAGGAGCGCTTGGGCGCAAGCTCCC 1200
DB 1141 CCTGAAGACGGCTTTCGCGGCGCGCGCATCAAGGAGCGCTTGGGCGCAAGCTCCC 1200

QY 1201 AAGCAGAGCGGCTATCTGTCAGCGCGGCTGCTGACCTTCTGCGGCGGCGGATAT 1260
DB 1201 AAGCAGAGCGGCTATCTGTCAGCGCGGCTGCTGACCTTCTGCGGCGGCGGATAT 1260

QY 1261 GACCGAATCATGAGCAGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCGCTGTGATG 1320
DB 1261 GACCGAATCATGAGCAGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCGCTGTGATG 1320

QY 1321 CCAGGACATCCCGGCGCACCCCGGTGAGCTGAGTCTGAGCGTCTATGTTTCAACCTCAAGCT 1380
DB 1321 CCAGGACATCCCGGCGCACCCCGGTGAGCTGAGTCTGAGCGTCTATGTTTCAACCTCAAGCT 1380

QY 1381 CCGGAGGCTATATGTTTCCAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 CCGGAGGCTATATGTTTCCAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

QY 1441 GTGAGACCGGAGGAAATATTACCCACCATCTACCGTGGTCAAGAGAGTGAATGAAC 1500
DB 1441 GTGAGACCGGAGGAAATATTACCCACCATCTACCGTGGTCAAGAGAGTGAATGAAC 1500

QY 1501 GGTCTCTATCTGCTTCCGATGATGATGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 1557
DB 1501 GGTCTCTATCTGCTTCCGATGATGATGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 1557

RESULT 2

AF467711
LOCUS 1560 bp mRNA linear PLN 01-JAN-2003
DEFINITION Emericella nidulans RfeC (rfeC) mRNA, complete cds.
ACCESSION AF467711
VERSION AF467711.1 GI:27450538

KEYWORDS

SOURCE Emericella nidulans
ORGANISM Emericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emericella.
1 (bases 1 to 1560)

REFERENCE

Holtzman, D.A.
Enhanced Production of Secondary Metabolites in Filamentous Fungi
by Ectopic Expression of Regulatory Proteins
Unpublished

JOURNAL

REFERENCE 2 (bases 1 to 1560)
Holtzman, D.A.
Direct Submission
Submitted (14-JAN-2002) Microbia, Inc., One Kendall Sq. Building
1400W, Cambridge, MA 02139, USA

FEATURES

source
1. .1560
/organism="Emericella nidulans"
/mol_type="mRNA"
/db_xref="taxon:162425"
1. .1560
gene

	Matches	1557;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0
Qy	1	ATGGATCTCTAGAAACCATCTCTGGCCCTCCGCTTACCAAGTCTGCTCAAGAGATCGCG	60							
Db	277	ATGGATCTCTAGAAACCATCTCTGGCCCTCCGCTTACCAAGTCTGCTCAAGAGATCGCG	336							
Qy	61	CTCTTCTCTTCTGCTCCCATCTCGAGCATGCGCAATGCTCTCAGTACACGATGAGCCTCAG	120							
Db	337	CTCTTCTCTTCTGCTCCCATCTCGAGCATGCGCAATGCTCTCAGTACACGATGAGCCTCAG	396							
Qy	121	TACCCAGTCTCTCAGCGCGACACCTTCTCTCTGCAACCCCATCATAGCCATGCGCC	180							
Db	397	TACCCAGTCTCTCAGCGCGACACCTTCTCTCTGCAACCCCATCATAGCCATGCGCC	456							
Qy	181	GCTCTCTACTCTGATCATGCGGCGAGCGCGTACCGGCTGATCTGAAACAGGTACCCCGCA	240							
Db	457	GCTCTCTACTCTGATCATGCGGCGAGCGCGTACCGGCTGATCTGAAACAGGTACCCCGCA	516							
Qy	241	TCAAGTCTCAGATGTTTACGGCTTCTGCTGCGCGCATATGCCCCCACTACCTGCGGC	300							
Db	517	TCAAGTCTCAGATGTTTACGGCTTCTGCTGCGCGCATATGCCCCCACTACCTGCGGC	576							
Qy	301	AGCTTGGCTTCCGACATCTTCTCTCTCATCCCAATCCGAGGCGCAGGCGACAGCGCAG	360							
Db	577	AGCTTGGCTTCCGACATCTTCTCTCTCATCCCAATCCGAGGCGCAGGCGACAGCGCAG	636							
Qy	361	CAATCGCGCATATCTCTCTCTCATACGCTGCTCCCGCCGCTTCCAGCGCTCAGTCTG	420							
Db	637	CAATCGCGCATATCTCTCTCTCATACGCTGCTCCCGCCGCTTCCAGCGCTCAGTCTG	696							
Qy	421	TACCGCGACCAATTGCGCGCGCGCTCCCGGAGCGTCTGCTGCTGACTTCAACAATGGA	480							
Db	697	TACCGCGACCAATTGCGCGCGCGCTCCCGGAGCGTCTGCTGCTGACTTCAACAATGGA	756							
Qy	481	CTTCTCTCAGGAGCATTCAGTTATTTCGACGGAAGCCTCAAGGTTGGGACCCCGTCT	540							
Db	757	CTTCTCTCAGGAGCATTCAGTTATTTCGACGGAAGCCTCAAGGTTGGGACCCCGTCT	816							
Qy	541	CGCAATGTGTCTGCGCGGTATCTCCGGGAAGGACTCCCCCGCAACCCAGGTTGTTGTTCT	600							
Db	817	CGCAATGTGTCTGCGCGGTATCTCCGGGAAGGACTCCCCCGCAACCCAGGTTGTTGTTCT	876							
Qy	601	CAGGGCGACCGGTATCTCTCGAGTGTTCGGGAGCGCGCAATCTCGGTCAACAATGGT	660							
Db	877	CAGGGCGACCGGTATCTCTCGAGTGTTCGGGAGCGCGCAATCTCGGTCAACAATGGT	936							
Qy	661	GTTAAGCGCACCGCAAGAACACTACTATCTCCGGCCAGGATGCGGATGGAAGTTCCT	720							
Db	937	GTTAAGCGCACCGCAAGAACACTACTATCTCCGGCCAGGATGCGGATGGAAGTTCCT	996							
Qy	721	TGCCGGAATCTGTAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC	780							
Db	997	TGCCGGAATCTGTAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC	1056							
Qy	781	CACACTGGTGACCGCGGTACATGTGTGTTTTCGAAAGACACTTCTCTCGAGTGAT	840							
Db	1057	CACACTGGTGACCGCGGTACATGTGTGTTTTCGAAAGACACTTCTCTCGAGTGAT	1116							
Qy	841	ATCTCTGAAACGTCTATTTCCAAAAATGCTCAATCAGGGCGTGAACCCACCGGAGCAACG	900							
Db	1117	ATCTCTGAAACGTCTATTTCCAAAAATGCTCAATCAGGGCGTGAACCCACCGGAGCAACG	1176							
Qy	901	CACTTGTGCAACCCCAATGGCGATGTGAAGAGGTCCCAACAGCAGGCTGGCGGAAATCTCT	960							
Db	1177	CACTTGTGCAACCCCAATGGCGATGTGAAGAGGTCCCAACAGCAGGCTGGCGGAAATCTCT	1236							
Qy	961	GTAACACTGTTCAGGATGAAGTCAGTGTACCGTCCCGCTCCCAATGSCATCCCGGGC	1020							
Db	1237	GTAACACTGTTCAGGATGAAGTCAGTGTACCGTCCCGCTCCCAATGSCATCCCGGGC	1296							
Qy	1021	ACGACTTACGCGGAGGAGCGCTCAACGGCAATGGACTAGCTCCGCGCCCGGCAGGTTAC	1080							
Db	1297	ACGACTTACGCGGAGGAGCGCTCAACGGCAATGGACTAGCTCCGCGCCCGGCAGGTTAC	1356							

```
Db 59 GCGAATGGAACCGCTAGAGACACACGATTCGGGCAAGATGCGGAGGCAAGTTCCT 118
Qy 721 TGCCGGAACCTGTAACAGACTTATCTTCATGCAAGCATCTCAAGCGCATCTGCTAGCG 780
Db 119 TGCCCTCATGTACCAAGACGACTCTCCAGCGGAGCATCTTAAAGCGCATCTTTTGA 178
Qy 781 C-----ACA 784
Db 179 CGTAAGTGTGTTCCACTCAATTCATTACCTTTTCCATCTATTAAACATCAAGCGCAG 238
Qy 785 CTGGTGACCGCGGTACATGCTGTTCTTTGCAAGACACCTTCTGCGAGTGATATCC 844
Db 239 CCGGTGATCGCCCTTACATGTGTCTCTGCAAGATATCTTCCGAGCGATATTT 298
Qy 845 TGAAGACGTCAATTTCCAAAATGCTCAATCAGCGGTGTAACCCCAACCGGAGCAACGCACT 904
Db 299 TGAAGACGACACTTCCAAAAGTCTGTTAGCGCGGTAAACCCACAGGAGCGAGCATC 358
Qy 905 TGTGCAACCCCAATGGCATGTGAGAGGTCCCAACAGAGCGGTGGGGGATCTGTAA 964
Db 359 TCTGCGATCTCAGGCGCACCTGAAACGGTC-----TCAGGCGCGAGCTAATGCTGCA 412
Qy 965 RACCTGTCCAGGATGAAGTCAGTAGTACGCTCCGCTCCCAATGSCATCCCGGCGACGA 1024
Db 413 AATCGGTTGAGGAAGTCAAGTAATCCATTCGCGCTTCCATGTTATGCGGTACGA 472
Qy 1025 CTTAGCGGAGGAGCGCTCAAGCGCAATGAGTACTGCTCCGGC-----CCGGCCAG 1075
Db 473 CTTATGGCGAGCGGACGCTGAATGGCATGGCCACAGCGTGGTCCGGTGGCGGACCTG 532
Qy 1076 GGTAGCGGATACAGACTATGGGCTCCCAATGTCATCGCTCAAGCGGATGGGCGGTG 1135
Db 533 GTCTCAGCGAGCAGCGCTCTGGGATATTCATGTTCTCTGCCCCAGCGGCTGAACAATG 592
Qy 1136 CTCAGCCTGAAGACGGTTTCGGCGGCGCGGCGCATCAAGGAGCGCCCTTGCCACAG 1195
Db 593 GTACGTTGAGACGGTTCAACCTCGCCAGGAGATCAGAGGTTCTGATGGGTG 652
Qy 1196 CTTCCAGCAGAGCGCGTATCTGTCAGCGGGTCTGACCGCTTCTGCCC 1246
Db 653 CTCAGAAGCAGAGCGCGTACTTGTCATCTTCTGTTCTGCTGAAGTTC 703

RESULT 6
LOCUS 166494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1. (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. G.
TITLE Recombinant fowpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
1. 7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 3.9%; Score 61; DB 6; Length 7218;
Best Local Similarity 1.9%; Pred. No. 0.00018;
Matches 7; Conservative 229; Mismatches 139; Indels 0; Gaps 0;
Qy 61 CTTCTTCTCTGCTCCATCTGAGATGCCAATGCTCAGTACAGGATGCGGCTCAG 120
Db 1070 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1129
Qy 121 TACCAGTCTCTCAGCGGACACCGCTCCTCTGCAACCCCATCATGACGAGTCGCC 180
```

```
Db 1130 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1189
Qy 181 GCTCCTCACTGATACATGGGCGAGCGCGCTACCGGCTGATCTGAACAGATGACCCCGCA 240
Db 1190 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1249
Qy 241 TCAAGTCAGATGTTTACGCGCTCTCTGTCGGCGGATAATGCCACACATACGCTGGGC 300
Db 1250 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1309
Qy 301 AGCTTGCTCGGACATCTTTCTCTCATCCCAATCGCAGGCGCAGCAGCGCAG 360
Db 1310 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1369
Qy 361 CAATGCGGCGACTATCTCTCTCATAGGTGTCGCGCGCGCTTCAGCGCTCAGTCG 420
Db 1370 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1429
Qy 421 TACCGCGAGCCAAAT 435
Db 1430 YYYYYGTACCAAT 1444

RESULT 7
LOCUS AC136406/c
DEFINITION Rattus norvegicus clone CH230-522A18, *** SEQUENCING IN PROGRESS
ACCESSION AC136406
VERSION AC136406.2 GI:25188350
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 241859)
Munzy,D.Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Gladly,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louised,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelimen,O., Okwouu,G., Olarumpasagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
```

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinkamp, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.

TITLE JOURNAL

REFERENCE AUTHORS

TITLE JOURNAL

REFERENCE AUTHORS

TITLE JOURNAL

COMMENT

2 (bases 1 to 241859)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (01-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 241859)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 23, 2002 this sequence version replaced gi:24462281.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KCQC
 Center clone name: CH230-522A18
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 187090 bases at least Q40
 Consensus quality: 189991 bases at least Q30
 Consensus quality: 191799 bases at least Q20
 Estimated insert size: 211470; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 5 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 64842: contig of 64842 bp in length
 64843 64942: gap of unknown length
 64943 70527: contig of 5585 bp in length
 70528 70627: gap of unknown length
 70628 237692: contig of 167065 bp in length
 237693 237792: gap of unknown length

* 237793 238825: contig of 1033 bp in length
 * 238826 238825: gap of unknown length
 * 238926 241859: contig of 2934 bp in length.

FEATURES

source

1. 241859

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-522A18"

2456..3351

/note="clone boundary"

clone end:Sp6

site:MboI

end sequence:BZ285495"

complement(52803..53490)

/note="clone boundary"

clone end:T7

site:MboI

end sequence:BZ285493"

102059..103179

/note="wgs end extension"

clone end:T7"

113425..117805

/note="wgs end extension"

clone end:T7"

BASE COUNT 51073 a 44181 c 45354 g 52522 t 48729 others

ORIGIN

Query Match 3.8%; Score 59.6; DB.2; Length 241859;

Best Local Similarity 45.9%; Pred. No. 0.00034;

Matches 203; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

QY 16 CATCCCTCTGGGCTTCGCTTACCAAGTCTGCCTCAAGATCGGCGCTTCCTTCTGCT 75

Db 87310 CCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 87251

QY 76 CCCATCTCTGAGCAGCCAAAGCCCTCAGTACAGATGAGCCCTCAGTACCCAGTCTCTCAG 135

Db 87250 CTGGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 87191

QY 136 CCGCACACCTCTGCTCTGCTCAACCCCATCATAGCAGTCCGCGCTCTCCTCCTGCTAC 195

Db 87190 CCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 87131

QY 196 ATGGGGGAGCGCGGTACCGGCTGTATCTGAACAGGTACCCCGCATCAAGTCAAGTGT 255

Db 87130 CCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 87071

QY 256 TAGCGCTCTCTGCTGCGCGGATATGCCCACTACCGTGGGAGGTGCTGCTCCGACA 315

Db 87070 CCTCTGCTCTGCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 87011

QY 316 TCTTTCCTTTCTCATCTCCCAATCCGAGCGCAGGCGCAGGCGCAGGCGCAGGCGCAGGCG 375

Db 87010 CCTCTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 86951

QY 376 CTTCTCTCTCATAGGCTGCTCCCGCGCTTCAGCGCTCAGTCTGCTACCCGAGCCCAATT 435

Db 86950 CCTCTGCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 86891

QY 436 GCGCGCGGCTCCCGGAGCC 457

Db 86890 CCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 86829

RESULT 8 AC083889/c

LOCUS DEFINITION

AC083889 214690 bp DNA linear HTG 05-OCT-2000

Mus musculus chromosome 5 clone RP23-271A20 strain C57BL6/J,

WORKING DRAFT SEQUENCE, 29 unordered pieces.

ACCESSION AC083889

AC083889.1 GI:10645305

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214690)
Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Dietrich, N.L., Gupta, J., Ho, S.-L., Idol, J.,
Lee-Lin, S.-Q., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B.,
Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B.,
Stantiriopop, S., Thomas, J.W., Thomas, P.J., Tiongon, E.E.,
Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A., Wetherby, K.D. and
Green, E.D.
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214690)
AUTHORS Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2000) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
COMMENT ----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.mouse@nih.gov
----- Project Information
Center project name: ry
Center clone name: 271A20
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 199272 bases at least Q40
Consensus quality: 203984 bases at least Q30
Consensus quality: 206372 bases at least Q20
Insert size: 184000; agarose-fp
Quality coverage: 4.91x in Q20 bases; agarose-fp
Quality coverage: 4.27x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2274: contig of 2274 bp in length
* 2275: gap of unknown length
* 2375: contig of 2235 bp in length
* 4610: gap of unknown length
* 4710: contig of 2084 bp in length
* 6793: gap of unknown length
* 6894: contig of 2144 bp in length
* 9037: gap of unknown length
* 9137: contig of 2568 bp in length
* 9138: contig of 2274 bp in length
* 11706: gap of unknown length
* 11806: contig of 2101 bp in length
* 13906: gap of unknown length
* 14006: contig of 3518 bp in length
* 14007: gap of unknown length
* 17524: contig of 982 bp in length
* 17625: gap of unknown length
* 22607: gap of unknown length
* 22707: contig of 5826 bp in length
* 28532: gap of unknown length
* 28533: contig of 6021 bp in length
* 34653: gap of unknown length
* 34754: contig of 6985 bp in length
* 41738: gap of unknown length
* 41839: contig of 7959 bp in length
* 49797: gap of unknown length
* 49898: contig of 6739 bp in length
* 56637: gap of unknown length
* 56736: contig of 5133 bp in length
* 61969: gap of unknown length
* 61970: contig of 7796 bp in length
* 69865: gap of unknown length
* 76103: contig of 6238 bp in length
* 76203: gap of unknown length
* 83336: contig of 7133 bp in length
* 83436: gap of unknown length
* 87846: contig of 4410 bp in length
* 87947: gap of unknown length
* 95043: contig of 7097 bp in length
* 95143: gap of unknown length
* 102460: contig of 7317 bp in length
* 102560: gap of unknown length
* 102561: contig of 6983 bp in length
* 109544: gap of unknown length
* 117915: contig of 8272 bp in length
* 118015: gap of unknown length
* 118016: contig of 9164 bp in length
* 127179: gap of unknown length
* 127180: contig of 12141 bp in length
* 134220: gap of unknown length
* 134221: contig of 9422 bp in length
* 148943: gap of unknown length
* 149043: contig of 13137 bp in length
* 162279: gap of unknown length
* 174494: contig of 15215 bp in length
* 177595: gap of unknown length
* 190534: contig of 12940 bp in length
* 190535: gap of unknown length
* 190635: contig of 24056 bp in length.
FEATURES
source
Location/Qualifiers
1..214690
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone_lib="RPCI mouse BAC library 23"
1..2274
misc_feature
/note="assembly_fragment"
2375..4609
misc_feature
/note="assembly_fragment"
4710..6793
misc_feature
/note="assembly_fragment"
6894..9037
misc_feature
/note="assembly_fragment"
9138..11705
misc_feature
/note="assembly_fragment"
11806..13906
misc_feature
/note="assembly_fragment"
14007..17524
misc_feature
/note="assembly_fragment"
17625..22606
misc_feature
/note="assembly_fragment"
22707..28532
misc_feature
/note="assembly_fragment"
28633..34653
misc_feature
/note="assembly_fragment"
34754..41738
misc_feature
/note="assembly_fragment"
41839..49797
misc_feature
/note="assembly_fragment"
49898..56636
misc_feature
/note="assembly_fragment"
56737..61869
misc_feature
/note="assembly_fragment"
61970..69765
misc_feature
/note="assembly_fragment"
69866..76103
misc_feature
/note="assembly_fragment"
76204..83336

```
misc_feature /note="assembly_fragment"
83437. .87846
/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature 87947..95043
/note="assembly_fragment"
95144..102460
/note="assembly_fragment"
102561..109543
/note="assembly_fragment"
109644..117915
/note="assembly_fragment"
118016..127179
/note="assembly_fragment"
127280..139420
/note="assembly_fragment"
139521..148942
/note="assembly_fragment"
149043..162179
/note="assembly_fragment"
162280..177494
/note="assembly_fragment"
177595..190534
/note="assembly_fragment"
190635..214690
/note="assembly_fragment"
BASE COUNT 55371 a 51412 c 50852 g 54227 t 2828 others
ORIGIN
Query Match 3.7%; Score 57.2; DB 2; Length 214690;
Best Local Similarity 45.7%; Pred. No. 0.0014;
Matches 200; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
Qy 16 CATCCCTCGCGCTCGCTACCACTGCTGCTCAGGATCGGCGCTCTCTCTGCT 75
Db 194143 CCTGCGCTGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 194084
Qy 76 CCCATCTGAGCATGCCAATGCCCTCAGTACGATGACGATGACGATGACGATG 135
Db 194083 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 194024
Qy 136 CGCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 195
Db 194023 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193964
Qy 196 ATGGGGCAGCGCGCTGACCGGCTGATGACGAGTACCGGCTGACGATGATG 255
Db 193963 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193904
Qy 256 TACGCGTCTTCTGCTGCGCGCATATGCCCCACACTACCGTGGGCGAGTGGC 315
Db 193903 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193844
Qy 316 TCTTCTCTTCTGATCCCAATCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAG 375
Db 193843 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193784
Qy 376 CCTCTCTCTATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 435
Db 193783 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193724
Qy 436 GCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 453
Db 193723 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193706
```

RESULT 9

```
AC087329
LOCUS AC087329 219559 bp DNA linear ROD 23-OCT-2002
DEFINITION Mus Musculus Strain C57BL6/J chromosome 5 BAC, RP23-383N15,
Complete Sequence, complete sequence.
ACCESSION AC087329
```

AC087329.10 GI:24270683

HTG.

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 219559)
Montgomery, K.T., Grills, G., Han, J., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
High Throughput Mouse Sequencing

TITLE

Unpublished

JOURNAL

2 (bases 1 to 219559)

REFERENCE

Montgomery, K.T., Grills, G., Han, J., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

AUTHORS

Direct Submission

TITLE

Submitted (28-DEC-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA

JOURNAL

3 (bases 1 to 219559)

REFERENCE

Montgomery, K.T., Grills, G., Han, J., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

AUTHORS

Direct Submission

TITLE

Submitted (29-JUN-2002) Harvard Partners Center for Genetics and
Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
02139, USA

JOURNAL

4 (bases 1 to 219559)

REFERENCE

Montgomery, K.T., Grills, G., Han, J., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,
Perera, A., Shim, C., Thomas, E., Brown, W.A. and Kucherlapati, R.

AUTHORS

Direct Submission

TITLE

Submitted (23-OCT-2002) Harvard Partners Center for Genetics and
Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
02139, USA

JOURNAL

On Oct 23, 2002 this sequence version replaced gi:21629249.

COMMENT

-----Genome Center:
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpcgg.org/Sequence/mouse.html>
Contact: hpgcmendel.mgh.harvard.edu

CLONE LENGTH: This sequence represents the entire insert of this
clone unless otherwise noted. If there are overlapping clones, the
overlaps are noted in the beginning and end of the features
listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550).

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.Genes and Regions of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
and cDNA sequences in Unigene. Genes demonstrate at least two exons
flanked by consensus splice sites that maintain sequence continuity
across the splice junctions. Sequences that are not identical
matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double
stranded sequence for all regions. All sequence is completed to a
standard of coverage with a minimum of 3 reads with no ambiguities.
If the sequence coverage for a region does not meet this standard,
it is indicated in the annotation as low coverage. Low coverage
linkages are verified by PCR product size verification or
verification of forward and reverse reads from clones which span
the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated average error rate is less than 1 per 10,000
bases using the Consed quality parameters. Regions that do not
meet this requirement are annotated as Low Quality.


```

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBQ2
Center clone name: CH230-6C1
----- Summary Statistics
Consensus quality: 220893 bases at least Q40
Consensus quality: 223841 bases at least Q30
Consensus quality: 225204 bases at least Q20
Estimated insert size: 223358; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 224928: contig of 224928 bp in length
* 224929 225028: gap of unknown length
* 225029 226314: contig of 1286 bp in length
* 226315 226414: gap of unknown length
* 226415 227533: contig of 1119 bp in length.
* Location/Qualifiers
*     1..227533
*     /organism="Rattus norvegicus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:10116"
*     /clone="CH230-6C1"
*
68901 a 53377 c 49887 g 54274 t 1094 others
                                     3.4%; Score 52.2; DB 2; Length 227533;
                                     1 Similarity 47.4%; Pred. No. 0.026;
                                     190; Conservative 0; Mismatches 208; Indels 3; Gaps 1;
160  GCCTCTTCTCTGCTCCCATCTCGAGCATGCCAAATGCCTCAGTACAGATGACGCTCA 119
111  GCCTCAGCTCAGCCCGCAGCTCAGCCCGCAGCTCAGCCTCAGCCCGCAGCTCAGCCTCA 177670
220  GTACCCAGTCTCTAGCCGCAACACCTTGCTCTCTGTGCAACCCCATCATGACGATGCGC 179
71  GCCTCAGGCCAGCTCAGCCCGCAGCTCAGCCTCAGCCCGCAGCTCAGCCCGCAGCCCA 177730
80  CGCTCTCACTCGTACATGGGCGAGCCGCTGACCGGCTGATCTGAACAGGTATCCCGC 239
331  GCCTCAGGCCCTAGCCCGCAGCTCAGCCTCAGCCTCAGCCTCAGCCTCAGCCCGCAGCTCA 177790
440  ATCAAGTCAACGATGTTACGGCTCTTCTGTGCGCCGATTAATGCCCCACATACCGTGG 299
91  GCCTCAGCTCAGCCCGCAGCTCAGCCTCAGCCCGCAGCTCAGCCTCAGCCTCAGCCCGCA 177850
00  CAGCTTGCTCGCAATCTTCTCTTCTATTCACATCCGAGCGCGCAGCGCA 359

```

Db 177851 GCCTCAGCCTCAGCCCCCAGCCCCCAGCCTCAGCCCCCAGCCTCAGCCCCCAGCCTCAGCCTCA 177910

Qy 360 GCAATCGCGCACTAT---CCTCCTCCTCATAGCGTGTCTCCGCGCGCTTCAGAGCGCTCA 416

Db 177911 GCCCAGCCCCAGCCTCAGCCTCAGCCTCAACCCAGCCCCCAGCCCCCAGCCTCAGCCTCA 177970

Qy 417 GTCGTACCGCAGCCAAATTGGCCGCGGCGCTCCCCGGGAGCC 457

Db 177971 GGCCAGCCTCAGCGTCAGCCTCAGCCCCCAGCCCCCAGCCCCC 178011

RESULT 12	AC122509	178757 bp DNA linear	ROD 22-MAY-2003
LOCUS	AC122509	Mus musculus chromosome 16 clone RP24-443G20,	complete sequence.
DEFINITION	AC122509		
ACCESSION	AC122509.2	GI:30985117	
VERSION	HTG.		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 178757)		
AUTHORS	Wilson, R.K.		
TITLE	The sequence of Mus musculus clone		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 178757)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	3 (bases 1 to 178757)		
AUTHORS	Wilson, R.K.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-MAY-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	On May 22, 2003 this sequence version replaced gi:21105973.		

FEATURES	source
	Location/Qualifiers
	1..178757
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="16"
	/clone="RP24-443G20"
BASE COUNT	52542 a 37167 c 36110 g 52938 t
ORIGIN	

[illegible][illegible]

RESULT 13.

AC128967	AC128967	197714 bp	DNA	linear	HTG 20-NOV-2002
LOCUS	Rattus norvegicus clone CH230-194J4	***, 3 unordered pieces.			SEQUENCING IN PROGRESS
DEFINITION					
ACCESSION	AC128967				
VERSION	AC128967.3	GI:25139175			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				

REFERENCE
AUTHORS

1 (bases 1 to 197714)

Muzny, D. Marie, Metzker, M. Lee., Abramzon, S., Anguiano, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Amigiano, D., Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. I., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulsegod, H., Lozado, R. J., Lui, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmood, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, D., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekemele, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K., Patel, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Guebragere, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Huiyk, S., Hume, J., Idelir, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lotenshuwa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaekelemeh, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Pankoch, C., Paeternak, S., Paul, H., Perez, A., Perez, L., Pflamkoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Stead, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vallas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wiczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

TITLE

REFERENCE

AUTHORS

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

Center project name: GKGW
Center clone name: CH230-219A19
Assembly program: Atlas 3.0;
Consensus quality: 235174 bases at least Q40
Consensus quality: 237474 bases at least Q30
Consensus quality: 238916 bases at least Q20
Estimated insert size: 248588; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 249856: contig of 249856 bp in length
249857 249956: gap of unknown length
249957 251187: contig of 1231 bp in length
251188 251287: gap of unknown length
251288 254531: contig of 3244 bp in length
254532 254631: gap of unknown length
254632 258694: contig of 4063 bp in length
258695 258795: gap of unknown length
258795 264103: contig of 5309 bp in length.

FEATURES
source
1. 264103
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-219A19"

misc_feature
1. 1292
/note="wgs contig"

misc_feature
31430..32987
/note="wgs contig"

misc_feature
246583..249856
/note="wgs contig"

BASE COUNT 66489 a 50509 c 53251 g 69852 t 24002 others

ORIGIN

Query Match 3.3%; Score 51.6; DB 2; Length 264103;
Best Local Similarity 44.7%; Pred. No. 0.036;
Matches 201; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 16 CATCCCTCTGGGCTCGCTTACCACTGCTGCTTCAAGGATCGGGCTCTTCTTCTGCT 75
Db 236448 CTTCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 236389

QY 76 CCATCTCGAGCATGCCAATGCTCAGTACAGATCAGATCAGATCAGATCAGATCAG 135
Db 236388 CTGCTCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 236329

QY 136 CGCACACCTGCTCTCTTCAACCCCAATATAGCAGTCCGCTCTCTCTCTCTCTCT 195
Db 236328 CTTCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 236269

QY 196 ATGGGGGACCGGCTTACCGGCTTATGTGAACAGTACCGGCTCTCTCTCTCTCT 255
Db 236268 CTGCTTCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 236209

QY 256 TACGGGCTCTTCTGCTGGCGGATATATGCCACACCTACCGTGGGACGCTTGGCT 315
Db 236208 CCACCTCCGCTTCCGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 236149

QY 316 TCTTCTCTTCTCTATCCCAATCCGAGGCGGACAGGACAGGACAGGACAGGACAG 375
Db 236148 CTTCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 236089

Center: Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

COMMENT

On May 13, 2003 this sequence version replaced gi:23113703.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2004, 19:09:19 ; Search time 3550 Seconds
(without alignment)

10659.738 Million cell updates/sec

Title: US-10-029-180-7

Perfect score: 1557

Sequence: 1 atggatcctagaacacatcc.....ccgttcagccgcagacaa 1557

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_esti:*

9: gb_est1:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_esthum:*

18: em_estinv:*

19: em_estpin:*

20: em_estvrt:*

21: em_estfun:*

22: em_estmam:*

23: em_estmus:*

24: em_estpro:*

25: em_estrod:*

26: em_estphg:*

27: em_estvrl:*

28: gb_est1:*

29: gb_est2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	380.4	24.4	442	9	AA785001
2	171.6	11.0	768	12	EM077105
3	108.8	7.0	687	10	BF251517
4	82.8	5.3	782	28	AQ161501

5	82	5.3	585	28	AQ255041
6	56.2	3.6	807	13	EX464554
7	53.8	3.5	1448	29	CC220110
8	52.6	3.4	1201	9	AL515475
9	52.2	3.4	727	13	EX319543
10	51.4	3.3	1137	29	CC315110
11	50.6	3.2	567	28	AZ895651
12	50	3.2	1201	13	EX457568
13	49.6	3.2	1201	13	EX421672
14	49.4	3.2	802	29	CNS01170
15	49.2	3.2	786	28	AQ745021
16	49.2	3.2	897	29	AG060530
17	49.2	3.2	915	13	BQ952432
18	49.2	3.2	1262	12	BM560943
19	49.2	3.2	1287	10	BF256624
20	49	3.1	481	9	AA695157
21	49	3.1	759	13	BU314584
22	48.8	3.1	834	28	BZ051898
23	48.8	3.1	1101	29	CNS00008
24	48.8	3.1	1201	9	AL514195
25	48.6	3.1	458	10	BF063108
26	48.6	3.1	648	12	BI227246
27	48.6	3.1	1026	29	AG081192
28	48.4	3.1	732	29	EX149729
29	48.4	3.1	815	29	EX239144
30	48.4	3.1	987	29	CNS00418
31	48.4	3.1	1137	29	AG078502
32	48.2	3.1	786	28	BH018846
33	48.2	3.1	2152	29	CC283954
34	48	3.1	785	29	AG081505
35	48	3.1	1130	29	AG060387
36	48	3.1	1201	13	EX356664
37	47.8	3.1	759	29	AG126157
38	47.8	3.1	774	29	AG128958
39	47.8	3.1	887	29	CNS028C6
40	47.8	3.1	1082	9	AL574528
41	47.8	3.1	1104	13	BQ671688
42	47.8	3.1	1122	29	CC231036
43	47.6	3.1	731	29	CNS04208
44	47.6	3.1	789	14	CA320647
45	47.6	3.1	1201	13	EX365323

ALIGNMENTS

RESULT 1
AA785001
LOCUS
DEFINITION

442 bp mRNA linear
g4b03al.r1 Aspergillus nidulans 24hr asexual developmental and
vegetative CDNA lambda zap library Emericella nidulans cDNA clone
g4b03al 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA785001
AA785001.1
GI:2845169

EST.
Emericella nidulans (anamorph: Aspergillus nidulans)

Emericella nidulans

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; Emericella.

1 (bases 1 to 442)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
Prade, R. and Roe, B.
An Aspergillus nidulans EST Database
Unpublished
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center

Applicants Copy

